The version of record is available at: http://dx.doi.org/10.1561/1500000073

Search Interface Design and Evaluation

Chang Liu[1], Ying-Hsang Liu[2], Jingjing Liu[3] and Ralf Bierig[4]

[1] *Peking University; imliuc@pku.edu.cn*
[2] *Oslo Metropolitan University; Ying-Hsang.Liu@oslomet.no*
[3] *University of Texas; jliujingjing@gmail.com*
[4] *Maynooth University; ralf.bierig@mu.ie*

ABSTRACT

This monograph reviews research on the design and evaluation of search user interfaces that has been published within the past 10 years. Our primary goal is to integrate state-of-the-art research in the areas of information seeking behavior, information retrieval, and human-computer interaction on the topic of search interface. Specifically, this monograph (1) describes the history and background of the development of the search interface; (2) introduces information search behavior models that help conceptualize users' information needs, and how people seek, select, and use information; (3) characterizes the major components of search interfaces that support different subprocesses based on Marchonini's information seeking process model; (4) reviews the design of search interfaces for different user groups, especially that of vulnerable people, as well as personalized and adaptive search interfaces; (5) identifies evaluation methods of search interfaces and how they were implemented in research having different evaluation purposes. We also provide an outlook on the future trends of search interfaces including

Chang Liu, Ying-Hsang Liu, Jingjing Liu and Ralf Bierig (2021), "Search Interface Design and Evaluation", Foundations and Trends® in Information Retrieval: Vol. 15, No. 3-4, pp 243–416. DOI: 10.1561/1500000073.
©2021 C. Liu et al.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

conversational search interfaces, search interfaces supporting serendipity and creativity, and searching in immersive and virtual reality environments.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

1

Introduction and Historical Background

Information seeking and use is now routine in people's everyday lives. Searching through various information retrieval (IR) systems such as web search engines or search functions within information systems allows users to gain access to information on the Internet. Whereas most research in this area has focused on the algorithms behind the search engines from technical perspectives, an aspect vital to system development, in this monograph, we focus on the search interface, the place where searchers interact with the search system. In some books and research papers, 'search user interface' is the term used to highlight the human users of search systems and to emphasize how the search interface should be designed to be appealing to a wide variety of people (Hearst, 2009). In the current monograph, the terms 'search user interface' and 'search interface' are used interchangeably.

The design of search interfaces has had a long history. According to Hearst (2009, p. 1), the search interface supports the four main tasks users carry out, 'expression of their information needs', 'formulation of their queries', 'understanding of their search results', and 'keeping track of the progress of their information seeking efforts.' The development of search interfaces and the mode of interaction between the user and

The version of record is available at: http://dx.doi.org/10.1561/1500000073

the search interface have changed with increasing velocity along a spectrum of trajectories. The interfaces of search systems have evolved dramatically with the development of human-computer interaction (HCI) technologies. Search systems have become ubiquitous with both oral and visual communication channels and capable of being conversational and intelligent (White, 2018). Search activities are often considered easy tasks for users, but increase in difficulty with more demanding types of search tasks. That is, fact-finding and navigational tasks are easier to accomplish than complex learning or exploratory tasks. The search interfaces ideally should be able to help users resolve a wide range of information problems in both their working and living environments, and support users in finishing the entire work task or achieve their information goals, not only support the search aspect. The design of search interfaces needs to consider users' complete search process, be informed by the theories and practices of user search behavior, and apply appropriate technologies to accommodate different groups of users in various contexts.

This monograph aims to present a comprehensive review of the design and evaluation of search user interfaces in the last decade. Since there are several comprehensive reviews of search user interfaces from 2009-2012, (*e.g.*, Hearst, 2009; Wilson, 2011; Russell-Rose and Tate, 2013), there is no need to go back further. In the past 10 years, studies in information science, IR, and HCI have had a better understanding of users' search interactions, including cognitive and behavioral mechanisms in the search process and the implementation of new technologies, such as automatic speech recognition, virtual reality (VR), and artificial intelligence (AI), to support informational activities and sensemaking. Through our review of recent contributions in related disciplines for the design of search interfaces, we hope to shed light on how to better apply the newly developed technologies to solve users' information problems in the workplace and in their everyday lives.

This section first presents a brief history of search interfaces; it then reviews previous review books and important review papers on search interfaces. The section closes with a description of the scope of this review and the structure of the following sections.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

1.1 The History of the Search Interfaces

Search interfaces are the place where users interact with search systems. However, the first large-scale operational IR systems were non-interactive (Cool and Belkin, 2011). For example, the Medical Literature Analysis and Retrieval System (MEDLARS), which was launched in 1964, only allowed the submission of requests to be queued at the National Library of Medicine for groups of searches against tapes. Thus, there was no interaction between end-users and the retrieval system at this stage. Real interaction did not happen until some sort of terminals were provided and users were connected to search systems online. For example, Medline replaced MEDLARS and began to provide search services to end-users in 1972. Later, there was a worldwide movement in libraries to replace card catalogs with online public access catalogs (OPACs). The retrieval systems and search interfaces were rooted in the field of Library and Information Science. The retrieval systems were designed to help users to retrieve documents from document collections or libraries, a task typically done by librarians. Many researchers agree that OPACs were the first type of end-user IIR systems (Savage-Knepshield and Belkin, 1999; Borlund, 2013). Therefore, the review of the early stage of search interfaces included both retrieval systems and OPACs.

This section reviews several interaction styles of search interfaces as they appeared chronologically in the early years before the modern search interfaces for web search engines had appeared. These four interaction styles are command-language interaction style, form fill-in interaction style, menu-driven interaction style, and direct manipulation interaction style (Borlund, 2013; Shneiderman *et al.*, 1997). For a list of the interactions they supported, see Table 1.1. The development of these four interaction styles demonstrated that end-users were being given more functions and options to interact with the search interfaces throughout the history of the design of search interfaces (Kelly, 2015).

1.1.1 Command–Language Interaction Style

At the beginning of the design of search systems, roughly from the mid-1960s to the mid-1970s, command–language interaction was the sole

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Table 1.1: Four interaction styles of search systems

Interaction style	Example search systems	New techniques to support users' interactions
Command-language interaction style	AIM/TWX, DIALOG, MEDLARS, NASA/RECON, the SMART system, The Biomedical Communication Network	• Display of online thesauri to help with query formulation; • Choice of novice or experienced searcher interface mode; • Ability to save search queries to rerun at a later time or on a different database; • Relevance feedback • System prompts for further information from user about his/her information need.
Form fill-in interaction style	THOMAS system	Adopted a cognitive viewpoint; engage users directly with texts; base user-system interaction around feedback
Menu-driven interaction style	RABBIT system	Provide selections from multiple commands
Direct manipulation interaction style	TileBars, book house fiction retrieval system	Provide visual representation

style of search interfaces due to the constraints of information technology. Command-language interfaces required searchers to construct search formulation phrases or sentences to search within an explicit framework of system files and commands. During this period, several operational IR systems were designed and developed, for example, AIM/TWX, DIALOG, MEDLARS, NASA/RECON, the SMART system, and the State University of New York (SUNY) Biomedical Communication Network (Walker, 1971).

Figure 1.1 is a sample dialogue from the AIM/TWX system, which shows the representation of the command-language interface during that period (Katter and McCARN, 1971). In this system, searchers could enter either a search statement or a command. In this example, the user first typed the command "aimlh", which invoked the display of an explanation of AIM. The user then entered another command "version

The version of record is available at: http://dx.doi.org/10.1561/1500000073

short all" to which the system responded by showing the abbreviations of all routine system messages. Next, the searcher entered "neighbor dopa", and the system responded with a list of the neighboring terms of the term "dopa" retrieved from the index.

```
...
USER: aimlh
...
PROG:
AIM-TWX:

THE ACRONYM "AIM" STANDS FOR ABRIDGED INDEX MEDICUS.
THIS IS A SUBSET OF INDEX MEDICUS WHICH INCLUDES CITATIONS
FROM THE ONE-HUNDRED ENGLISH LANGUAGE BIOMEDICAL JOUR-
...
SS 1/C? - - - SEARCH STATEMENT 1 OR COMMAND? - -
ENTER SEARCH STATEMENT NUMBER 1 OR ANY COMMAND.

USER: "version short all"

PROG:
SS 1 /C? - - - SEARCH STATEMENT 1 OR COMMAND?

USER: "neighbor dopa"

PROG:
POSTINGS    TERM

1       DOORY Y (AU)
2       DOOUSS TW (AU)
- -     DOPA (MH)
...
```

Figure 1.1: Sample dialogue from AIM/TWX

The search interface for the DIALOG system was a question-answer negotiation process. It provided a command input function for well-defined information needs and also provided a browsing function (the so-called "expand" function on the interface). After clicking this function, the interface showed terms that were alphabetically near to the search term in the query with the intent to help searchers better understand their information needs. The NASA/RECON system also provided an "expand" function for searchers and showed the thesaurus structure of related terms in the query.

During that time, Boolean operators were adopted in the retrieval algorithms, but some search interfaces, like the SUNY system, concealed the use of the underlying Boolean expressions by asking in a prompt window, "Do you want to add another subject to this group?" This was

The version of record is available at: http://dx.doi.org/10.1561/1500000073

the first implementation of this kind in search interface design that did not force searchers to formulate a command as a query.

Even though the IR was only able to support searchers with the specific information they wanted, researchers realized searchers' queries, especially the original queries, were often inadequate. Researchers desired to know more of users' interactions with the IR systems, but in the meantime, suggested librarians "show searchers a few books in an attempt to pinpoint searchers' needs" (Ide, 1967; Ide, 1969).

By the middle of the 1960s, several interface techniques had been introduced to assist end-users (Kelly, 2015), including:

- Displaying online thesauri to help with query formulation (*e.g.*, the DIALOG system and the NASA/RECON system);
- Providing a choice of novice or experienced searcher interface mode (*e.g.*, the DIALOG system);
- Concealing the use of Boolean expressions (*i.e.*, AND, OR, NOT) during query formulation by prompting users with questions, such as "Do you want to add another subject to this group?";
- Enabling the saving of search queries to be rerun at a later time or on a different database (*e.g.*, the SMART system);
- Providing relevance feedback (*e.g.*, the SMART system);
- Adding system prompts for further information from the user about his/her information needs (*e.g.*, the SUNY Biomedical Communication Network).

In 1971, the first workshop about interactive search interfaces was held (Bennett and Walker, 1971). In this workshop, Bennett presented his challenge paper, proposing questions on how to design search interfaces to support various levels of user expertise, the conceptual framework of the appropriate level of interactions that search interfaces should support, and how to evaluate search interfaces and IR systems from the users' perspective. Bennett's design challenges continue to guide and influence research and practice in user-system interaction to

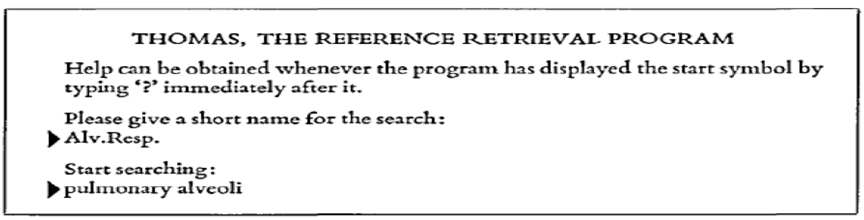
The version of record is available at: http://dx.doi.org/10.1561/1500000073

this day, and have led to substantial progress in the development of search interface design.

1.1.2 Form Fill-in Interaction Style

From the mid-1970s to the mid-1980s, designers of search interfaces believed that a reference retrieval system should aim to “help the user to make choices from among unseen documents” (Oddy, 1977). Most of these retrieval systems were designed to target novice searchers (Savage-Knepshield and Belkin, 1999). It was during this period that the form fill-in interaction type emerged, the THOMAS retrieval system being one such example. Users could interact with the system by inputting simple statements through dialogues. During this stage, IR was completed through a man-machine dialogue. An example retrieval process is shown in Figure 1.2 to Figure 1.5.

The THOMAS system was one of the first experimental IR systems that adopted a cognitive viewpoint in its design. Searchers could engage in a dialogue about their ill-defined information problem using this system. THOMAS is notable for being the first interactive IR system to engage users directly by way of texts and to base user–system interaction around feedback.


THOMAS, THE REFERENCE RETRIEVAL PROGRAM

Help can be obtained whenever the program has displayed the start symbol by typing ‘?’ immediately after it.

Please give a short name for the search:
▶ Alv.Resp.

Start searching:
▶ pulmonary alveoli

Figure 1.2: THOMAS system Homepage and an example first query

1.1.3 Menu-Driven Interaction Style

Form fill-in interaction style required searchers to understand field labels and know the permissible values for the fields. Comparatively, another style, the menu-driven interaction style, was more appropriate for novice

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Influence of fasting on blood gas tension, pH, and related values in dogs.; Pickrell *et al, Am J Vet Res, 34,* 805–8, Jun 73
1. J A Pickrell, 2. J L Mauderly, 3. B A Muggenburg, 4. U C Luft, 5. animal experiments, 6. animal feed, 7. arteries, 8. blood, 9. body temperature, 10. carbon dioxide, 11. dogs, 12. fasting, 13. hemoglobin, 14. hydrogen-ion concentration, 15. irrigation, 16. lung, 17. oxygen, 18. pulmonary alveoli, 19. respiration, 20. time factors
▶?

Figure 1.3: The reference presumed to be of the most interest to the searcher is shown, together with a series of associated terms or author names

There can be three parts to your statement (all optional):
1. Your reaction to the reference just shown (if any). This must come first:
   "Yes" or "No"
2. A selection from the names (authors) or terms shown, by number. A "not" in the statement signifies rejection of all numbers that follow it.
3. New names or terms (terms preferably in quotes). The elements of the statement should be separated by commas.

Examples: 'posture', 'circulatory system'
Yes, not 11,12
No, 7,13,4
'heart rate'
Yes

Press enter key when you are ready to proceed ▶

Figure 1.4: Assistance interface available upon searchers' request

searchers, providing searchers with a limited number of options to choose from during their search process. The RABBIT system (Tou *et al.*, 1982) is an example of this type. As shown in Figure 1.6, after entering a query, the searcher could enter attribute values. In response to the query, the system displayed one example instance from the database in detail along with a menu containing all other matches. To refine a query, the searcher would select an attribute to modify his query and then choose from five commands displayed in a pop-up menu in a context-sensitive manner as appropriate for that specific attribute. The provision of the labels on the menus of the search interface helped significantly in reducing the users' cognitive load by swapping recall memory tasks with recognition tasks from a list of options so that searchers could focus more on their

The version of record is available at: http://dx.doi.org/10.1561/1500000073

... 12. fasting, 13. hemoglobin, 14. hydrogen-ion concentration, 15. irrigation, 16. lung, 17. oxygen, 18. pulmonary alveoli, 19. respiration, 20. time factors
?
▶No, 10,17,19,20

Figure 1.5: The searcher's sample reply to the dialog after he is done with the instruction

searching tasks (Shneiderman, 1983).

1.1.4 Direct Manipulation Interaction Style

The direct manipulation interaction style (Shneiderman, 1983) was implemented by a hypertext approach characteristic of the Berry-picking model (Bates, 1989; Bates, 1990). This, coupled with the advent of the graphical user interface (GUI), provided more flexibility and control for end-users during their search resulting in the use of retrieval systems by more and more untrained novices. A wealth of research examined the effects of the individual characteristics on users' search performance and search interactions in a quest to learn how to design IR systems that could better accommodate individual differences through interactions and search interfaces.

The appearance of the GUI near the end of the 1980s have made search interfaces more interactive since that time. The BookHouse fiction retrieval system designed by Pejtersen (1989) was an icon-based retrieval system designed to support casual novice users in their search for fiction books. On the homepage, the searcher was presented with a picture of a house built of books, a visualization of the public library environment (Figure 1.7). The left room had books for children, the right room had books for adults while the center room had books for both. The direct manipulation interaction features allowed the user to click directly on the figures executing different strategies as he/she usually did in a physical library. Novice searchers were able to self-explore the search system without extra training.

This brief review of the history of search interfaces demonstrates

The version of record is available at: http://dx.doi.org/10.1561/1500000073

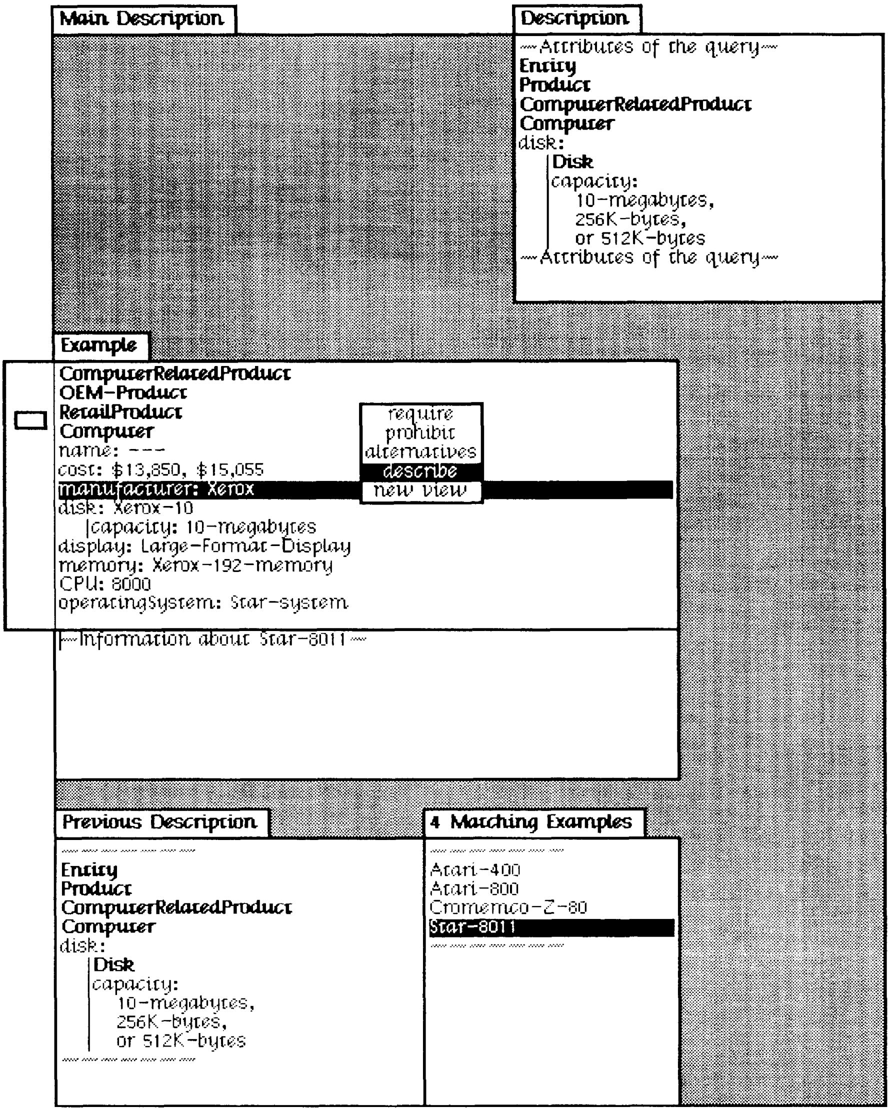


Figure 1. RABBIT Screen Display

Figure 1.6: A screenshot of the RABBIT system

The version of record is available at: http://dx.doi.org/10.1561/1500000073

that search systems have experienced a movement from a technology-dominated stage to a user interactive mode. The GUI and other display techniques in HCI have advanced the research and implementation of search interfaces of retrieval systems.

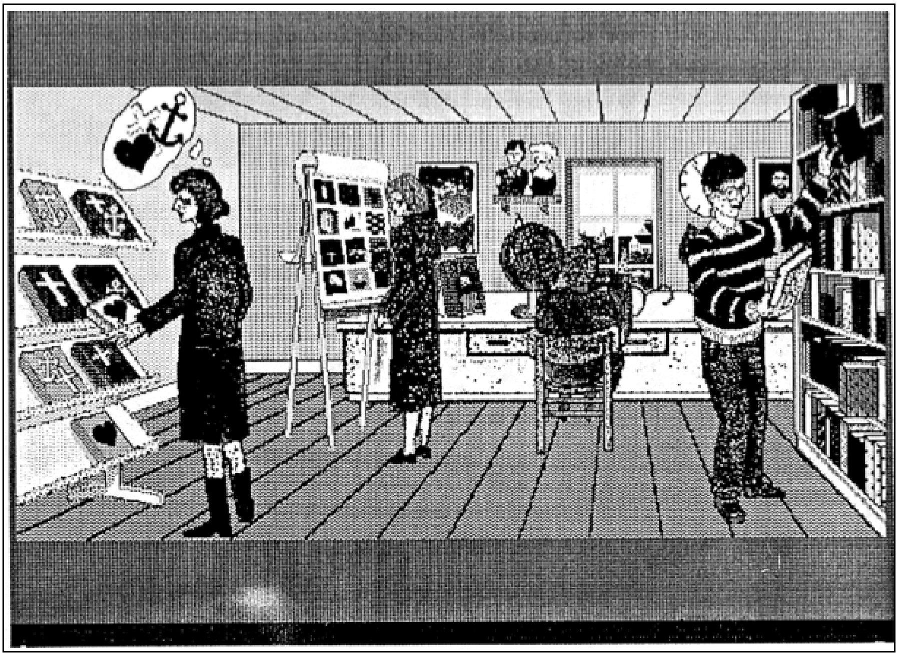


Figure 1.7: One of the search interfaces of the BookHouse fiction retrieval system

1.2 Previous Review on Search Interfaces

Ever since the first workshop on search interfaces was held in 1971, "The User Interface for Interactive Search of Bibliographic Data Bases" (Bennett and Walker, 1971), great progress has been made in the development of search interfaces resulting in them being more effective and efficient for end-users. In this workshop, Bennett and Walker (1971) were the earliest in paying serious attention to the interactive properties of IR and proposed a set of design challenges to researchers in the field. Of the several important review books and articles on search interfaces in the 1990s, the review written by Savage-Knepshield and Belkin (1999)

The version of record is available at: http://dx.doi.org/10.1561/1500000073

took the 'Bennett challenge' as a guiding framework and reviewed the historical development of search interfaces from the 1960s to the end of the 1990s.

Search User Interfaces (Hearst, 2009) was the first academic book to focus on the search user interface. It provided a comprehensive review of the human side of the information seeking process, described the methods for search interfaces design and evaluation, and discussed research results surrounding various components of search interfaces, (*i.e.*, query specification and query reformulation, the display of search results, grouping retrieval results, navigation of information collections, search personalization, and the broader tasks of sensemaking and text analysis). Max L. Wilson (2011) wrote a review shortly after that in 2011 highlighting the more complicated and exploratory scenarios that led people to search and to evaluate whether their search was successful. In this book, Wilson reviewed a large number of search user interface features and designs, and explored how they could support searchers with different kinds of intentions. The search features that Wilson reviewed were classified into four categories: input features, control features, informational features, and personalizable features. Russell-Rose and Tate (2013) published their book from the information architecture perspective, in which they reviewed theories in information seeking and wove that with the practice of search user interface design. They applied the principles of user-centered design not only to the search box and to the display of search results, but also extended it to faceted navigation, mobile interface, social search, and so on, and on multiple devices, such as desktop, tablet, mobile, and others.

In this decade, we have seen the widespread usage of search services by online searchers in more complicated and exploratory scenarios, accessing more diverse online resources and websites, and being initiated from various interactive devices. Besides the comprehensive review of search interfaces, there have also been several review books on specific topics of search, for example, faceted search by Tunkelang (2009). Faceted search has been prevalent in online information access systems, particularly for e-commerce and site search. Tunkelang (2009), in his review of its history, theory, and practice, states that faceted search is based on the faceted classification of information, which could also be a

The version of record is available at: http://dx.doi.org/10.1561/1500000073

fundamental theory of knowledge organization in all kinds of representation and discovery tools (Broughton, 2017). In addition, since working in collaboration to perform information-seeking tasks has become more and more common, Hansen *et al.* (2015) provided a collection of best practices and studies in the field of collaborative IR and search.

In 2017, Ryan White (2016) published his comprehensive review book, *Interactions With Search Systems*, which summarizes the current state of many empirical studies on search interactions, but is not particularly about search interfaces. He also cast an eye toward the future of search systems forecasting that the next generation search systems will go beyond the query-response paradigm and will provide more reactive, proactive, and iterative experiences to searchers given the advances in technologies such as speech recognition and computer vision, new interaction capabilities such as touch and gesture, the emergence of cloud computing, and the democratization of AI. As these technologies will also be sure to influence the future development of search interfaces, we think it is timely and necessary to provide an update on the subject of search interfaces, in particular, one that focuses on the recent developments and new applications of the past 10 years.

1.3 Scope

Since both Hearst (2009) and Wilson (2011) have provided extensive reviews on how users search and interact with search systems and the design of search systems before the year 2010, the current review will focus, in particular, on recent developments and new applications of search interfaces in the past decade. Search interface design is an interdisciplinary field which involves information-seeking behavior research in information science, IR in computer science, HCI, and human-centered computing. We will try to include the research from all of the above areas and other related areas as well that focus on how to implement search interfaces 1) for more complicated and exploratory searches, 2) in different domains and for different groups of users, and 3) with the help of the advances in new technologies, as well as 4) how to evaluate users' experience with search interfaces.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

The structure of this review is as follows: Section 2 provides related theories and models in information seeking and search behaviors, and more importantly, it includes recent discussions on the application of work tasks and search tasks in search interface design. These theoretical developments help us build the framework on which to support users' search processes through search interfaces. Section 3 then explains how search interface features are designed to support different search processes, namely, the searching process, the browsing and selection process, and the process of working with the information. Section 4 begins to consider search interface design for different groups of people, for various domains, and on different devices, issues which have not been fully reviewed in previous review books since these advances have been recent, mainly occurring within the past 10 years. Section 5 details the methods for evaluating search interfaces including evaluation approaches, evaluation measures, and other concerns. The last section, Section 6, discusses the search interfaces of next-generation search systems which may incorporate and implement more advanced technologies, for example, physiological signal-based search interfaces, gaze-based search interfaces, gesture-based search interfaces, adaptive interfaces, conversational interfaces, and searching in immersive and VR environments.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

2

Information Search Behavior Models

Designing successful search interfaces requires an understanding of peoples' needs, how they seek for information, and how they will use the information they collect during the search process. Research in information seeking and search behaviors has proposed many theories and models and conducted empirical studies to better understand human information seeking tasks and strategies, all of which have implications for designing search systems that can accommodate users' tasks and search strategies, as well as for evaluating retrieval systems in the context of these tasks and strategies. This section reviews a wide range of highly-cited information seeking and search behavior models. Given the limited space, it is difficult to include all the significant models and theories in information seeking and search behavior. Readers who are interested in delving deeper into these models could refer to Fisher *et al.* (2005). The first section of this monograph reviews nine important search behavior models, and the second section describes the concept of tasks and their application in the design of a search interface. The section then concludes with a discussion of the application of information search behavior models in search interface design.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

2.1 Search Behavior Models

This section selects nine popular and influential models that are frequently referenced in interactive information retrieval (IIR) and widely used by researchers and practitioners in the design of search interfaces. The nine models reviewed cover different aspects of information search behaviors; specifically, Belkin's ASK model which describes users' knowledge state before and during searching; two models that highlight the sequentially evolving process of an information search: Bates' berrypicking model and Pirolli's information foraging theory; three descriptive process models: Ellis' information-seeking behavior model, Kuhlthau's information search process (ISP) model, and Marchionini's information seeking process model; two models about the representation of search interactions or information objects: Belkin's episode model and Ingwersen's cognitive model and polyrepresentation; and one integrative model for relevance: Saracevic's stratified model. Another dimension from which to think about these models is the task dimension: some tasks are at the work task level, for example, Kuhlthau's and Ellis's models; while others, like Marchionini's model, are at the search task level. We do not distinguish them in this section, but Section 2.2 will mainly discuss the application of tasks in search interface design.

2.1.1 Belkin's ASK Model

Belkin proposed the concept of the anomalous state of knowledge (ASK) within an explicitly communicative analysis of the fundamental problem of information science as "the effective communication of desired information between the human generator and human user" (Belkin, 1977, p.22). The ASK concept was proposed at the cognitive level. Within this model, texts are considered as representations of the conceptual states of knowledge of their generators and potential recipients. It perceives the reason why people look for information as being that the recipient has recognized an anomalous state of knowledge concerning some goal and desires to resolve the anomaly. To date, most approaches in the IR field had assumed that people knew what information they were seeking; however, Belkin claimed that most of the

The version of record is available at: http://dx.doi.org/10.1561/1500000073

time, people do not know what they do not know and cannot articulate the information they need. Therefore, he recommended that the retrieval system should not assume that the search process begins when people can express their information need or queries clearly or specifically.

The ASK model bears similarities to other models in information science, such as Taylor's first stage of the development of a need, which he describes as "visceral" or "unconscious", when an individual feels a "vague sort of dissatisfaction... [that is] inexpressible in linguistic terms" (Taylor, 1967, p. 182). Despite this similarity, ASK was produced to support the user-centered design of IR systems (Belkin, 1980a; Belkin *et al.*, 1982). As Belkin (1980b) pointed out, under the ASK hypothesis, it is inappropriate to ask a person to specify that which is required to resolve an ASK. Instead, the ASK should be represented in ways appropriate for representing what a person does not know and maybe discovered through dialog rather than specification. Belkin (1980a) elaborated on these ideas and suggesting how the ASK hypothesis could be implemented in IR systems.

Belkin (1980a) recommended that the seeking process should be viewed at its earliest inception, at ASK. The seeking process would then begin when a person realized ASK in their brain but were as yet unable to spell out their information need. Recently, with the development and implementation of functional magnetic resonance imaging technology in IR research, researchers have explored ways of designing search systems that could proactively satisfy searchers' information needs. Moshfeghi *et al.* (2019) examined users' brain signals during their process of analyzing a question to identify whether the user had recognized an ASK (which would provoke search systems with their information needs) or had had a successful memory retrieval (which would not result in an information seeking process).

Belkin (1980a) also presented ASK as an associative structure of concepts and their relationships; that is, as a graph or network. Subsequently, Belkin *et al.* (1982) suggested five classes of ASKs based on several characteristics of the graphs, for example, how well-formed and understood the ASK was by the user. Then Yuan *et al.* (2002) related ASK type to Relevance criteria categories and found that the structure of a person's information need (ASK) is related to the relevant criteria

The version of record is available at: http://dx.doi.org/10.1561/1500000073

which that person would bring to bear when evaluating documents. The ASK hypothesis also implies that the search system could help people articulate their information needs more clearly either through dynamic search sessions or through the elicitation of richer queries and background knowledge (Kelly and Fu, 2007).

2.1.2 Bates' Berry-Picking Model

The typical model in IR is a simple linear model, which does not capture the nature of the information search process by human beings. Bates (1989) took the metaphor of berry picking to describe the nature of users' search processes. A user's search intention may change by following up various leads and shifts in thinking. Therefore, the information search process proceeds bit by bit.

The berry-picking model of information search emphasizes three points in the user's information search process. First, usually, information search has an evolving query which shifts according to the feedback received during the search; second, the information search process proceeds bit by bit, like picking berries in a forest. Therefore, there need not be a single complete final retrieved set for any search; thirdly, users often use many different sources and search techniques throughout the course of searching. Bates (1989) highlighted browsing, in particular. She further described the process of browsing (Bates, 2007) as containing four elements, iterated indefinitely until the overall episode ends: (1) glimpsing a field of vision; (2) selecting or sampling a physical or representational object from the field; (3) examining the object; and (4) physically or conceptually either acquiring the examined object or abandoning it. She also gave some suggestions on how to design an interface to facilitate browsing.

For example, a good browsable interface would consist of rich scenes full of potential objects of interest that the eye could take in at once (massively parallel processing) and then select items from within the scene to which to give closer attention. One direction suggested by Bates (2016) as deserving of further exploration is "Design for real browsing", with special emphasis on the capability to scan. First, the screen needs to be large enough to allow the eye to take a glimpse of the data on a

The version of record is available at: http://dx.doi.org/10.1561/1500000073

page rich in content, especially in that it should have many points of potential interest to the user. There should be different types of search capability scattered on the screen, for example, taxonomies, which would support the hierarchical browsing of topics linked to content and facilitate more accuracy in searching. Faceted search, which divides search results into logic groups based on information architecture, is found to make the user's browsing experience more comfortable. Faceted navigation allows the user to elaborate his queries progressively, and facet search combines faceted navigation with text search so as to allow users to access semi-structured content collections (Tunkelang, 2009).

2.1.3 Pirolli's Information Foraging Model

In an analogy with the food foraging behavior of living organisms, information foraging theory (Pirolli and Card, 1999) is a general model describing how people use different strategies and technologies to locate information in response to the changing environment. The theory posits that one uses proximal cues to identify important content for further exploration or consumption. Based upon information foraging theory, one's choice of links is determined by the perceived cost and value of accessing various sources from proximal cues (Card *et al.*, 2001). Information foraging theory seeks to predict how rational information seekers behave when finding relevant information, and assumes that people often maximize information gain by minimizing the cost of information seeking. Further, they proposed three information models for information foraging theory: information patch model, information diet model, and information scent model.

The information patch model aims to predict the amount of time an information forager would spend within an information patch before searching for new patches. The information diet model assumes that information predators could be of different types; for example, a generalized information predator would pursue a wide range of relevant information with diverse dimensions while a specialized information predator would only collect from only a few relevant information sources having precise characteristics. The information scent model explains how people identify the value of information based on information cues,

The version of record is available at: http://dx.doi.org/10.1561/1500000073

for example, the result clusters shown on search engine result pages (SERPs), to gain an overall sense of the information space.

As high-level human cognition such as complex problem-solving is context-dependent Ragni (2020) information foraging theory has been applied to study search behavior to see if information searchers would alter their web search behaviors when SERPs were intentionally manipulated. For example, Wu and Kelly (2014) found that stronger information scent increased document examinations leading the searcher to click deeper into the search results on the desktop. However, Ong (2017) found that, on mobile, increasing the number of relevant search results beyond the initial screen size reduced the number of documents examined. Liu *et al.* (2010) proposed a user classification model containing three criteria: information goal (I), search strategy (S), and evaluation threshold (E). It is named the ISE model based on information foraging theory. Then they identified six user characters in the ISE model: fixed, evolving, cautious, risky, weak, and precise. They extracted users' interaction features from an image search experiment and performed multiple linear regression models of the three evaluations. These regression models and qualitative data analysis verified their ISE user classification model. Conceptualizing an information foraging scenario as sequential decision making, Wittek *et al.* (2016) examined user eye gaze behavior in information seeking by taking measurements of risk (hesitation) and ambiguity (opportunity cost) in an uncertain environment. The results revealed that users with different cognitive styles use different search strategies when performing a search in an uncertain environment.

2.1.4 Ellis' Information Seeking behavior Model

Ellis (1989) conducted a series of studies and examined the information-seeking patterns and characteristics of academic social scientists, research physicists and chemists, research scientists, and engineers in an industrial firm. He proposed and elaborated a general model of information seeking behaviors using the 'grounded theory' approach, and derived eight major characteristics of search seeking behaviors: starting, chaining, browsing, differentiating, monitoring, extracting, verifying,

The version of record is available at: http://dx.doi.org/10.1561/1500000073

and ending; he suggested implementing some system functions in a hypertext environment for each type of search behavior.

- Starting: initiating a search for information
- Chaining: following chains of citations or other forms of referential connection between materials or sources
- Browsing: performing a semi-goal-oriented search by browsing in an area of potential interests to find something of particular interest
- Differentiating: judging information sources based on type, quality, importance, or usefulness to his or her information need
- Monitoring: searching for information but for current awareness purposes where the user maintains an awareness of developments and technologies in a field
- Extracting: working through a particular source to locate material of interest in the extracting mode
- Verifying: checking information concerning correctness and reliability
- Ending: concluding the search, linking new information with previous knowledge

Among these "characteristics", Ellis (1989) did not suggest any particular order but suggested organizing the characteristics sequentially using logic. Wilson (1999), however, noted that "starting" should be the first stage and "ending" the last; browsing, chaining, and monitoring are search procedures, whereas differentiating is a filtering process, and extracting is an action performed on the information sources.

The identification of these categories of information-seeking behavior suggests that information search systems could include features or functions that directly support these activities to increase the usability and usefulness of the systems. For example, for starting, an individual could begin surfing the Web or start searching by identifying sources of

The version of record is available at: http://dx.doi.org/10.1561/1500000073

interests, or websites containing or pointing to information of interest or some popular authors; for chaining, they could follow links in the search result list or hyper-links within their selected pages that lead to other content-related sites; for browsing, they could scan the web pages of the sources selected or of the links on SERPs; for differentiating, they could bookmark some web pages for a later comparison of their content and usefulness; for monitoring, they could either revisit favorite sites for new information or receive site updates using push or agents; for extracting, they could highlight or extract content of interest on the web pages.

Ellis's 1989 empirically-based model of common information-seeking actions associated with scholarly information seeking has been influential such that follow-up studies were conducted to test for similar activities in the working circumstances in other domains (Meho and Tibbo, 2003), and specific functions of search systems and search interface features were proposed based on these actions.

2.1.5 Kuhlthau's Information Search Process (ISP) Model

Kuhlthau (1991) conducted a series of user studies on students in an educational setting to examine how the information seeking process develops throughout their whole process of doing a term project. After observing and recording students' information behaviors, cognitive stages, and affective aspects during the inquiry process, Kuhlthau proposed the ISP model which consists of six stages: initiation, selection, exploration, formulation, collection, and presentation. Common patterns of actions, thoughts, and feelings were found at each stage. According to Kuhlthau, although this sequence of tasks may appear somewhat recursive, the general process proceeds from the initiation to the completion of the project. From among these stages, Kuhlthau highlighted the exploration stage, which is often the most difficult stage for users because it is at this stage that users often have difficulties expressing precisely what information they need, and feelings of confusion, uncertainty, and doubt are frequently at their peak. However, the designers of search systems have often been unaware of it. As Kuhlthau's model highlights the affective dimension of users' emotions as they interact with information,

The version of record is available at: http://dx.doi.org/10.1561/1500000073

it has design implications for search systems. Building on Kuhlthau's work, Kalbach (2006) outlined a framework for understanding users' emotional states as they seek information on the Web. Kuhlthau's ISP model also indicates that people seek information to create, learn, and innovate in the context of their daily lives, so the design of the search system needs to accommodate users beyond searching.

Concerning the implication of the ISP model to the design of search systems, Russell-Rose and Tate (2013) compared these stages to a funnel that begins open-ended and ends with a decisive resolution and found most search applications invest their efforts mainly toward the end of the funnel: the collection and presentation. Many researchers have suggested extending the search functions and interface to support open-ended tasks, or so-called exploratory search, in modern search systems (White and Roth, 2009). As Marchionini (2006) claimed, for exploratory search, the search systems should not only provide look-up functions but also support learning and investigation, concepts similar to the early stages in Kuhlthau's ISP model.

As Huurdeman and Kamps (2015) has stated, current search systems mainly support cycles of micro-level interactions (*e.g.*, entering queries, selecting items from the results list), but do not explicitly support the macro-level information seeking as described in Kuhlthau's ISP model. Huurdeman *et al.* (2016) further suggested two ways to design a potential stage-aware system to support a user's information-seeking process on a macro-level. The first way is to allow the user to manually input which "stage" he is currently at and to select the appropriate interface to be shown, and the second way is to design the search system such that it could automatically detect the stage the user is in. Russell-Rose and Tate (2013) proposed three methods for assisting the users through this macro-level process: facilitate open-ended exploration with flexible browsing and filtering controls; help users organize their findings both to keep track of what they have encountered along the way and to monitor for new opportunities that sometimes arises. In recent years, the problem of designing knowledge-context into search systems to facilitate users' learning and critical thinking has attracted extensive research attention. For example, Azpiazu *et al.* (2017) introduced 'YouUnderstood.Me' (YUM), an evolving online environment built around a search engine

The version of record is available at: http://dx.doi.org/10.1561/1500000073

that can not only retrieve materials that satisfy information needs but also match the user's reading ability, thus making the search results valuable to children (5-15 year olds). Smith and Rieh (2019) proposed that SERPs should make both bibliographic and inferential knowledge context readily accessible to motivate and facilitate information-literate actions (such as comparing, evaluating, and differentiating between information sources) so as to support the metacognitive skills required for long-term learning, creativity, and critical thinking.

Vakkari condensed Kuhlthau's model into three problem stages, pre-focus, formulation, and post-focus, from his series of studies (Vakkari, 2000a; Vakkari, 2000b; Vakkari and Hakala, 2000; Vakkari and Pennanen, 2001). In his longitudinal study, students were asked to search for information when writing research proposals for a master's thesis to examine how the stages of writing a research proposal were related to the types of information searches, to the search tactics and term choices, and to judgments of relevance. Specifically, the pre-focus stage contains initiation, selection, and exploration stages in Kuhlthau's model; the formulation stage is the same as that in Kuhlthau's model, and the post-focus stage refers to collection and presentation stages.

Huurdeman *et al.* (2016) designed two three-stage tasks based on Vakkari's three problem stages to investigate the utility of search user interface features. They found that different search features were useful at different stages; for example, the informational features (search results) were always useful in all three stages, while the usefulness of input (search box) and control features (category filters, tag cloud, query suggestions) showed a downward tendency after the pre-focus stage; users preferred to use personalizable features (recent queries, saved results) after the pre-focus stage.

Gaikwad and Hoeber (2019) employed Vakkari's three-stage model of information seeking as a design guide in the context of interactive image retrieval for the image retrieval system, ImgSEE, and as a mechanism for controlling the laboratory-based evaluation. The study showed participants were able to follow the three stages of information seeking and make use of most of the features provided in support of each stage. They also observed that some participants seemed to return to the focus formulation stage after entering the post-focus stage, indicating that the search process may be more cyclical than linear.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

2.1.6 Marchionini's Information Seeking Process Model

Marchionini (1995)'s model is composed of a set of subprocesses (see Figure 2.1). Information seeking begins with the recognition and acceptance of the problem. Users would then choose a search system based on their previous experience with the task domain. Here, users would formulate a query, execute a query, examine results, extract information, and finally reflect or stop searching.

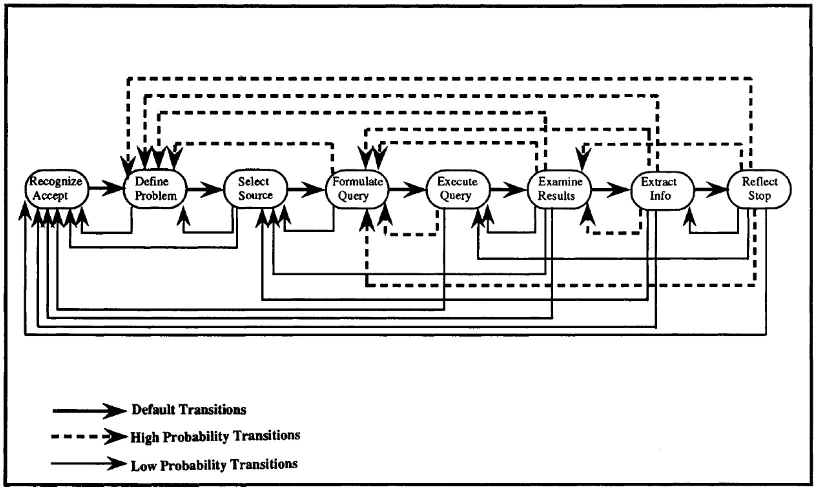


Figure 2.1: Marchionini's model of information seeking process within search contexts (Marchionini, 1995, p. 50).

Marchionini's information seeking process model conveys the eight key stages a searcher often progresses through during a search episode, and emphasizes the iterative nature of the process. The model also shows the transitions and their probabilities between these steps. Jürgens *et al.* (2014) interviewed several patent searchers and described the patent retrieval process on the basis of Marchionini's information seeking process model. Their alternation of the model was to neglect Execute Query, and to extend the Extract Info step to Extract Info/Report. To emphasize the importance of the contextual factor experience at all steps, they drew an ellipse surrounding the whole process.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Sahib *et al.* (2012) compared the visually impaired with sighted users during complex information-seeking tasks to better understand the challenges they met at different stages of the information seeking process based on Marchionini's model. They found that only very limited support was offered for query formulation, such as query suggestions and spelling support. It took longer for the visually impaired to explore search results due to the higher cost in time and effort. With respect to the reformulation stage, visually impaired users usually created new queries rather than reformulating queries due to the higher costs of the reformulation process. It was found that note-taking is generally poorly integrated with search engines.

The strength of the information seeking process model is that it is easier to move on to the next step, that of search system improvement, by addressing the challenges observed in information seeking scenarios (Sahib *et al.*, 2012; Marchionini and White, 2009). Even though the back-arrows shown in the model indicate that it is not normal to go only straight ahead in a prescribed order, that other paths are also possible, Wilson (2017) exclaims that the information search process should not be simplified into being represented as a linear left-to-right progression, but should be represented as a temporal progression. Based on this consideration, Wilson proposed the Tetris model as an analogy of the process of how searchers resolve their information needs.

Both Kuhlthau (1991) and Marchionini (1995) presented information seeking process models that operate from the temporal perspective, but each approach the search from different levels. Recently, Savolainen (2018) reviewed information seeking process models from the perspective of their temporal development, and summarized these models as being of two levels: the stage-based approaches and the cyclic models. The stage-based approach, (*e.g.*, Kuhlthau (1991)'s ISP model, Vakkari and Hakala (2000) and Xie (2009)) were proposed and research conducted on systems based on these models describe the different stages searchers often experience during a longitudinal work task or learning-related task. These studies found that the search stage played an important role in affecting users' search behaviors. The cyclic models, (*e.g.*, Marchionini (1995)'s model and research by Gwizdka (2010), Kules and Capra (2011), and Yue *et al.* (2014)) examined the search process of one search

The version of record is available at: http://dx.doi.org/10.1561/1500000073

session to describe how users switch among different sub-processes. Both approaches have advantages: stage-based approaches provide an overview of information-seeking processes usually spanning a longer period of time; cyclic models break the strict linearity and suggest individual constituents can reoccur in varying order or simultaneously within one or two stages. Savolainen (2018) further suggested that more effort in devising an integrated temporal framework covering both perspectives is needed to design search systems that will support the complex nature of the information-seeking process.

In the following section, we will also adopt Marchionini (1995)'s ISP model to demonstrate how the components in search systems were designed and implemented to support users' search processes.

2.1.7 Belkin's Episode Model

Belkin (1996)'s episode model views a user's interaction with an IR system as a sequence of episodes of different kinds. The term "episode" can be considered synonymous with "event" or "interaction". Traditional IR typically assumes users and their search goals are "static" and that a search could be done within the context of a single query-response cycle. However, this is not the case for real users. A user engages over time in several different interactions, each dependent on several factors, such as the user's current tasks, goals, and intentions, and the history of the episodes. Different kinds of interaction exist because they support a variety of processes, such as judging, interpreting, modifying, and browsing. This model defines the typical steps of interactions between a user and an information search system as "scripts", and highlights four binary dimensions (method, goal, mode, and resource) to define 16 unique information-seeking strategies (ISSs). Later, Belkin and Cool (2002) extended and expanded upon the four dimensions in the episodic model to incorporate all possible interactions among people and information within five facets using a faceted classification scheme.

This model could be served as the basis for the implementation of support techniques for different ISSs in search systems. Yuan and Belkin (2007) implemented an integrated IIR system that had been adapted to support different ISSs (both scanning and searching) and compared it

The version of record is available at: http://dx.doi.org/10.1561/1500000073

against a standard baseline search system. The results show that the integrated search system achieved better search performance and search experience. Huvila (2010) adopted Belkin and Cool (2002)'s classification scheme and extended it to the broad scope of information activity which included their contexts, and also highlighted the importance of maintaining a balance between complexity and simplicity when classifying information interactions. As Wilson *et al.* (2009) pointed out, this model may not be sufficiently exhaustive, representing only the core of search interactions, but further research using such classification could examine the relationship between search context or search intent and users' search interactions (*e.g.* Xie, 2000).

2.1.8 Ingwersen's Cognitive Model and Polyrepresentation

Ingwersen (1992), Ingwersen (1996), and Ingwersen (1999) developed and enhanced the cognitive model of IR interaction, and Ingwersen and Ingwersen and Järvelin (2006) proposed an integrated information-seeking and retrieval (IS&R) research framework based on the holistic cognitive viewpoint and relevant theoretical and empirical research in IS&R. This framework understands the IS&R as a process of cognition for the information-seeking actor(s) or team in context. Information seeking involves cognitive and emotional representations from a variety of participating actors. Such representations are seen as manifestations of human cognition, reflection, emotion, or ideas forming part of the IS&R components and kinds of interactions in context.

Typical information-seeking behavior is the acquisition of information from knowledge sources; interactive IR involves information acquisition via formal channels. Information acquisition, use, and interaction are thus regarded as central phenomena of information behavior, including IS&R. Every information actor operates in, and is influenced by, a dual contextual frame: that of the IT and information spaces surrounding the actor(s)—the systemic context and the socio-cultural organizational context. During the interaction, any actor is influenced by his past experiences (the historic context) and in turn, actors influence their systemic and socio-cultural environment, directly and indirectly, over time via other actors' information seeking and use of information.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Whereas the actor's cognitive and emotional states occur before and during each interaction, information systems are designed for all interactions, and their design is supposed to be stable for some time. To keep the symmetry, the system would have to choose an algorithm for each stage in a query. This capability has not yet come to be and is possibly an undesirable and unachievable goal.

Ingwersen's cognitive model indicates that information objects should be represented in different forms. Ingwersen (1994) developed the theory of polyrepresentation based on the observation that a combination of different types of representations of the same information objects tends to produce better IR results than when only a single type of representation is used. In practice, a combination of overlapping representations with different sources and functions (*e.g.*, author assigned article titles and indexer assigned metadata) would be a sign of greater relevance of an information object. The theory of polyrepresentation has been tested in and supported by several studies (*e.g.* Skov *et al.*, 2008; Schaer *et al.*, 2012; Huvila, 2016).

2.1.9 Saracevic's Stratified Model

The term "interaction" is a broad and general term. What is interaction? Saracevic (1997) proposed the stratified model to highlight that we can consider IR interaction as occurring on several connected levels or strata. IR interactions can be considered as a dialog between a user and a computer, occurring in episodes involving different levels or strata, and all these interactions happen on the search interface. Both the user side and the computer side have multiple factors that could influence the design of the search interface. From the user side, the user's tasks, intent, and knowledge structure could all have an impact on the user's query characteristics; from the 'computer side, the search interface reacts to the user's input based on text representation provided by the algorithms and hardware.

On the user side, there are three levels/strata: cognitive, affective, and situational. On the cognitive level, users interact with texts and their representations, (both can be considered cognitive structures); on the affective level, users interact with their intentions, such as beliefs,

The version of record is available at: http://dx.doi.org/10.1561/1500000073

motivation, and feelings; on the situational level, users interact with the given situation or problems-at-hand, and the results of the search may be applied to the resolution of the problem.

On the computer side, there are at least three levels as well: engineering, processing, and content. The engineering level involves hardware and its various operational attributes; the processing level concentrates on software, algorithms, and the various approaches of texts and queries; the content level concentrates on information resources and their various representations. The computer side of the model has been extended by other researchers. For example, Spink *et al.* (1998) added a graduated relevance dimension; and Bates (2002) identified additional levels that interact and affect each other.

User and computer sides meet at the surface level via an interface. For users, despite the complexity of search systems, the user and the system meet on the interface level interacting mainly through queries and the search results displayed. Empirical studies that have taken the stratified model as a theoretical framework have mainly examined users' query formulation and reformulation behaviors (*e.g.* Rieh and Xie, 2006). The main idea of the stratified model is to provide a holistic view of the interactions from both the user and computer sides. All the layers impact the performance of the search system and the users' search experience. Even though an effective search algorithm is implemented on well indexed data and efficient hardware, a non-user-friendly search interface design would hinder users' search experience and search performance. The designers of a search system could take the stratified model and check the features of each layer to identify possible bottlenecks hindering the performance (White, 2016).

2.2 Application of Search Tasks

In traditional IR research, an IR system was often thought of as a generic search system that would respond to a query with a set of results to meet some information need, whereas nowadays, the task is usually considered to be what triggered the need to search in the retrieval system (Toms, 2011). It is somewhat true that search can be considered a "solved problem" for fact-finding and navigational

The version of record is available at: http://dx.doi.org/10.1561/1500000073

searches, but the interaction model and the underlying algorithms are still brittle in the face of complex tasks. As White claimed, "...we need to invest in evolving search interaction to, among other things, address a broader range of requests, embrace new technologies, and support the often underserved "last mile" in search interaction: task completion." Search log analysis has shown that long search sessions are very common and that tasks often extend over long periods on and on more than one device (Hassan *et al.*, 2014). It is argued that search systems should be designed and evaluated based on their ability to assist users in accomplishing their higher-level tasks. In information seeking research, various tasks have been discussed, such as work tasks, information-seeking tasks, and information search tasks (Byström and Hansen, 2005). Among these types of tasks, the work task is often viewed as the motivation for other types of tasks. Work task refers to an activity people perform to fulfill the responsibilities of their job, such as a work-related task (Li and Belkin, 2010). Work tasks may not only refer to work assigned by others but could also refer to self-motivated tasks, like travel planning. In general work, tasks are situations wherein users have specific goals in mind that they need to accomplish, and that often involve searching. There are also situations in which users do not have specific goals, where they search for leisure or serendipity, or just search casually for information or browsing. Such searching will be further discussed in Section 6.3.

There have been several workshops focused on the task, including Larsen *et al.* (2012)'s Task-Based and Aggregated Search Workshop, the Second Strategic Workshop on Information Retrieval in Lorne (SWIRL), (Allan *et al.*, 2012) and Kelly *et al.* (2013)'s workshop on task-based information search systems. A commonly seen basis of this stream of research is to classify tasks into different types along with some task feature(s). These include, for example, closed versus open-ended tasks (Marchionini, 1989); factual, descriptive, instrumental, and exploratory tasks (Kim, 2006); fact-finding vs. information gathering (Kellar *et al.*, 2007; Toms *et al.*, 2007); learning about a topic, making a decision, finding out how to, finding facts, and finding a solution (Freund, 2008). The two most comprehensive examinations are those of Kim and Soergel (2006) and Li and Belkin (2008). Some of the task

The version of record is available at: http://dx.doi.org/10.1561/1500000073

attributes they developed include sources, time, product, process, goal, the complexity of tasks, features of the users, and the performance of the task by the user from the perspective of the user. In recent years, Search as Learning has attracted considerable research attention, and researchers have often adopted Anderson *et al.* (2001)'s five levels of cognitive complexity and the knowledge dimension as the design guideline for learning-related search tasks (Wu *et al.*, 2012; Urgo *et al.*, 2019).

Freund and Wildemuth (2014) created the Repository of Assigned Search Tasks (RepAST), which contains bibliographic details for empirical studies that have employed assigned search tasks as well as conceptual papers on task-based searching. When available, the search task types, definitions, and task descriptions themselves were included. RepAST provides a platform that contains search task types and task descriptions, through which to study the practices within the research community and prompt greater conceptual clarity and consensus in the use of search tasks. RepAST can be used as a source of search task descriptions for reuse in new studies or to replicate prior research.

In IIR research, studies are often conducted to understand how task characteristics (complexity, difficulty, etc.) or task stage influence users' search behaviors, users' judgments of document relevance, and search performance/outcomes. A deeper understanding of these relationships would help determine which task characteristics have actual design implications for the search system (for a search engine's ranking algorithm and/or the interface). From the system design perspective, it requires modeling and tracking a user's task over multiple queries, search sessions, and devices, and designing interactions that guide the user toward task completion, as well as developing evaluation methodologies that more directly measure a system's ability to help users complete the task at hand. Research efforts have been devoted to determining what types of search tasks a user is trying to accomplish (Mitsui *et al.*, 2018a; Mitsui *et al.*, 2018b); what kind of experience a user has had (*i.e.*, task difficulty, search frustration, search satisfaction, etc.) during searching for a task (Liu *et al.*, 2014; Chen *et al.*, 2015; Zhang *et al.*, 2018; Chen *et al.*, 2017); what categories best represent users' search intentions and how such intentions could affect the type of information users

The version of record is available at: http://dx.doi.org/10.1561/1500000073

expect (Mitsui *et al.*, 2017); how to implement task type information in the optimization of search results (Liu *et al.*, 2012); what search user interface features should be provided to support users at different stages of the search process (Huurdeman *et al.*, 2016). Yet more research is needed to determine what kinds of information about a user's task a search engine should try to predict; what kinds of information the search engine should elicit from the user directly; what are the appropriate mechanisms for eliciting task information; and what are the appropriate times to elicit information.

Users make use of a wide range of tools to accomplish their tasks. Huurdeman *et al.* (2016) designed two three-stage tasks based on Vakkari's three problem stages to investigate the utility of search user interface features based on Wilson (2011)'s classification of search features at different stages. They found that the informational features (search results) were always useful at all three stages while the usefulness of input (search box) and control features (category filters, tag cloud, query suggestions) showed a downward tendency in usage after the pre-focus stage. On the contrary, users preferred to use personalizable features (recent queries, saved results) after the pre-focus stage. Huurdeman (2017) later further proposed a theoretical framework for designing search user interfaces with enhanced support for macro-level processes. It outlines how three types of search user interface features (input and control features, informational features, and personalizable features) can be recombined to form a supportive framework for complex tasks. In this framework, Huurdeman suggested designing a stage-aware search user interface which would provide low-level support for moves and tactics, gradually giving way to higher-level support for stratagems and strategies. Besides supporting users' searching process, future work should also consider developing auxiliary tools that help users integrate and make better use of the information found during their searches.

2.3 Application of Models

As discussed in this section, there has been a disconnect between IR system design and models of user search behavior for multiple reasons (Fidel, 2012b; Kuhlthau, 2005). Wilson *et al.* (2009) proposed an

The version of record is available at: http://dx.doi.org/10.1561/1500000073

evaluation framework and evaluated advanced search user interfaces using search behavior models, specifically Belkin and Cool's ISSs and Bates' search tactics. Dillon (2016) proposed that a theory is needed for the design of search interfaces that support reading since it requires different kinds of theoretical knowledge for the design of interactive search systems.

Sonnenwald *et al.* (2016) identify types of theories developed in information sciences where Kuhlthau's ISP model has been further developed for explaining and predicting. Even though this model has implications for IR system design, few studies have directly applied it to the design of IR systems. Nonetheless, the ISP model has been translated into Web design practices by considering the user's emotional states (Kalbach, 2006) and has been extensively discussed as foundational in the design of search experiences (Russell-Rose and Tate, 2013). Huurdeman *et al.* (2016) found that there are specific relationships between search interface features and the stages of information-seeking tasks, that is, some features are more useful than others at a certain stage of information seeking. Sarraf (2019) identified the mapping of neural activities to the ISP model. Overall, these studies suggest that Kuhlthau's ISP model has been extended in theory to make connections with the usefulness of some system features as well as the user's neural activities. From a practical perspective, the ISP model has informed the design of search experiences and Web design practices by taking into account the user's emotional states.

The information foraging theory and associated concepts have been used to explain user search behavior in the desktop and mobile search environments (*e.g.*, Wu and Kelly, 2014; Ong, 2017). Informed by the information foraging theory, Schnabel *et al.* (2019) found that users prefer interfaces that have a lower access cost, regardless of the strength of information scents in recommended systems. Montoya Freire *et al.* (2019) applied the information foraging theory to the design of layouts in user interfaces. Generally, the implications of the information foraging theory for search interface design in these studies have not been adopted and tested in operational systems. In other words, the theory has been applied to explain user search behavior and provide implications for search user interface design. From practical perspectives, Russell-Rose

The version of record is available at: http://dx.doi.org/10.1561/1500000073

and Tate (2013) listed three techniques for implementing information scent into the design of the search user interface: descriptive titles, hit highlighting, and clear labeling. Given the development of Kuhlthau's ISP model and the information foraging theory, we need to further develop theories for the design of search user interfaces drawing from our understanding of the theories and models of user search behavior.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

3

Interfaces to Support the Search Process

Researchers and practitioners usually consider at least two broad areas relevant to the design of search interfaces: the design of the search feature (including its presentation and its interaction design) and the presentation of the search results for users to consume. The interactive process that users follow, however, can be separated into additional stages: formulating queries, examining search results, browsing, selecting and interacting with sources for further use, then finally completing the search task. Good search interfaces guide and assist searchers throughout this entire process from the initial idea to the completion of the search activity. In this section, we adopt Marchionini's information-seeking process model as a framework to review some of the studies and search interface designs concerning the different search process stages (Marchionini, 1995).

As Marchionini noted, the information-seeking process is dynamic and action-oriented, and these subprocesses can be depicted as belonging to one of four classes: understanding and planning, searching and execution, evaluation, and use, as shown in Figure 3.1). Among these four classes of action, the understanding and planning subprocesses are mainly mental activities, and the execution, evaluation, and use sub-

The version of record is available at: http://dx.doi.org/10.1561/1500000073

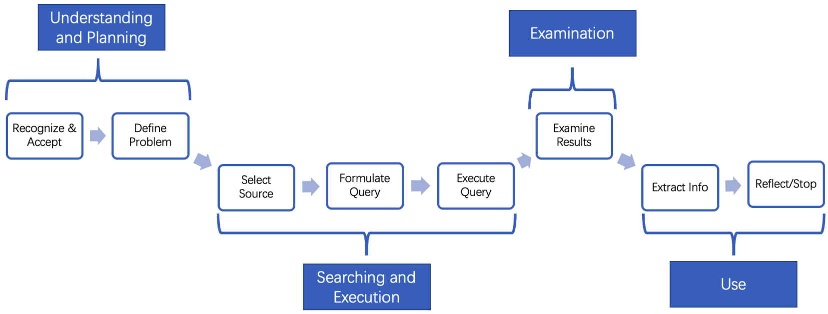


Figure 3.1: Structure of Section 3: The four subprocesess in Marchionini's model

processes are both mental and behavioral. In the past decade, various search functions or interaction styles have been designed to support different subprocesses. Search interface design should not only support searchers' behavioral subprocesses but should also support searchers' mental subprocesses. In the following sections, we will discuss how the search interface could support these four main subprocesses.

3.1 The Understanding and Planning Subprocesses

The first two steps in Marchionini's model are 'recognize and accept and information problem' and 'define and understand the problem', which have often been considered cognitive processes that only occur in searchers' minds, invisible to the search systems. However, recent research in IIR often argues that search systems should be designed to support searchers' work task completion (White, 2018). In real-world settings, searching is never an independent and isolated activity existing in a vacuum; instead, it is surrounded by other activities. Searching is usually motivated by some sort of task called a "work task" in the literature of IIR (Järvelin *et al.*, 2015; Li and Belkin, 2008). During the task completion process, searching is often accompanied by making annotations, taking notes, or copying and pasting information chucks, and so on; in addition, searching generally leads to knowledge gain (*e.g.*, Liu *et al.*, 2013), and sensemaking and learning mainly occur outside of the search engine (Wilson *et al.*, 2012a). Searching is

The version of record is available at: http://dx.doi.org/10.1561/1500000073

also generally followed by the use of the found information to achieve the work task goals. Information use is clearly different but closely connected with information search. Information use, in addition to information search, are both components of Wilson (1999)'s model of information-seeking behavior. Effectively using the information to accomplish a specific purpose is one of the standards in information literacy, alongside accessing and evaluating information, et cetera (ALA, 2016). Kuhlthau (1991)'s Information Seeking Process (ISP) model includes a "presentation" phase, when the search is complete and the problem has been solved, which also features information use. Supporting information use appears to be a naturally extended goal that search systems can reach.

To design a search interface able to support the understanding and planning subprocesses, it is important to know how searchers would use the information and the information outcomes they would produce after task completion. The use of information can be in the mental or cognitive format and generate mental outcomes such as thoughts and ideas, but it often results in physical and concrete products. Järvelin *et al.* (2015, p. 21) highlight the importance of the work task outcome, noting that it is not enough in a typical complex task scenario to just learn about a topic or to solve a problem with selected information items, but "the outcome of cognitive processing of information has to be documented and justified using information items as evidence." Järvelin *et al.* (2015) further propose that the ultimate goal of information interaction tools is to support the performance of work tasks. Supporting the accomplishment of work tasks, particularly, the work task products or outcomes (these two terms are used interchangeably in this section), has been attracting increasing research attention in recent years. Fourney (2015) points out that when relying on web resources to support their work, users interact with systems/objects in three environments: (1) the search engine, (2) retrieved documents, and (3) the application's user interface. Designing interfaces that can support work tasks is a significant aspect of support which requires first understanding people's information work tasks and planning goals.

In one such effort to understand the users' tasks and goals, Vakkari and Huuskonen (2012) examined how a client information system fared

The version of record is available at: http://dx.doi.org/10.1561/1500000073

in supporting social workers' tasks in child protection, specifically, how information production and use were embedded with other tasks in the whole work process. The client information system was found insufficient in its support of social workers' tasks due to problems such as there being too little information in the client information system to use, manual conversion of client records being required, the complexity of the system, etc. Recommendations included redesigning the system to be simpler, easier to use, and offering ways for users to find summarized information more rapidly. Although search is involved, the client information system is not an information search system like search engines or library databases that the current paper focuses on. However, this does give us an example of how to go about researching the need.

There have been tools designed to help with work tasks. MacKay and Watters (2008) designed browser tools to help users organize the large amount of information that they had saved and used in multi-session tasks. The tools also had features that could help users resume tasks from where they had left off when they returned to the tasks. These tools are browser add-ons and do not belong to search systems themselves, but the ideas may have implications for search system design.

Besides the tools to support information use and information production, more effort should be put into helping searchers better recognize and define their information problems or work tasks. As Kuhlthau (1991)'s ISP model emphasized, searchers often experience the exploration process to narrow down their general information need to a more focused topic for their work tasks, and such an exploration process is often associated with frustration, doubt, and uncertainty. Therefore, a call for new interactive types of search systems that could help searchers better understand their information problems and that could better support searchers' self-awareness and meta-cognition is needed.

3.2 The Searching and Execution Subprocesses

The searching and execution subprocesses, which have been the main focus in the design of search systems, mainly involve selecting search systems, formulating queries, and executing queries. In this section, we

The version of record is available at: http://dx.doi.org/10.1561/1500000073

review three aspects of support for one of these subprocesses, that of users' query formulation: search box design, query auto-completion to support query formulation, and content-based multimedia search.

3.2.1 Search Box Design

The most important rule in search box design is to make it noticeable and prominent. The search box should not be designed behind an icon; otherwise, users would have to take an extra action to access the search box. Besides having a prominent search box, it is also recommended that the search box be accompanied by a magnifying glass icon, the most universal search symbol recognized by users. Not all search systems prefer a single search box, take for instance the library search and casual search in the Digital Cultural Heritage (Walsh and Hall, 2015)

The most common form of text search on the Web is through keyword searching. A keyword is an index entry that identifies a specific record or document. In the early days of search systems, the author of the web document specified the keywords for the documents. Current search engines establish rules to extract and index words that appear to be important in the documents; for example, words found in the title of a page, or words that are repeated several times throughout the document. Keyword searching requires that the user type the same words the search engines generated from the documents for use as index terms, or the search engines need some other way in which to understand the searcher's intent and the contextual meaning of terms in the user's queries and within the documents.

Early query suggestion functions were derived from relevance feedback (Koenemann and Belkin, 1996; White *et al.*, 2005). For the next phase, the search system was designed to suggest queries from query logs (Baeza-Yates *et al.*, 2004). For query suggestions, there are two options, either to suggest the complete queries or to suggest individual terms, and researchers have found that searchers preferred complete queries be suggested rather than individual terms (Kelly *et al.*, 2009). In another follow up study, Kelly *et al.* (2009) compared the sources of the suggested queries and found that searchers preferred query suggestions from human-generated queries over suggestions generated by

The version of record is available at: http://dx.doi.org/10.1561/1500000073

algorithmic methods. Niu and Kelly (2014) examined the usage of query suggestions over successive queries within search sessions and found that the usage rates for the first eight queries of sessions were consistently about 50% and the usage was reduced for subsequent queries. These early forms of query suggestions are called static suggestions (Smith *et al.*, 2017).

3.2.2 Query Auto-Completion to Support Query Formulation

Besides static suggestions, there are dynamic query suggestions which enable assistance on the first query of a session. The first use of dynamic suggestions in commercial Web search engines was developed by Yahoo in 2007 (Anick and Kantamneni, 2008), followed closely by Google in 2008 (Smith *et al.*, 2017). One technique used in assisting users' query formulation dynamically is the query auto-completion (QAC) technique. QAC is the ubiquitous information search function that displays a list of suggested queries, where the list changes as the searcher types (Smith *et al.*, 2017).

The concept of auto-completion for queries is derived from typing prediction functions (*e.g.*, Jakobsson, 1986). Cai and Rijke (2016) have presented a detailed treatment of the history and development of QAC mechanisms. They mined query logs and used the popularity of queries to predict the complete queries. They added prefix and click-through features (Bast and Weber, 2006), the use of context such as the date of the query to make seasonal adjustments (*e.g.*, moving "Halloween" higher in the list during October), or a searcher's location, demographics, and individual search history (Shokouhi, 2013), trends in recent query activity at the search engine level (Whiting and Jose, 2014), and semantic and syntactic information found in large query logs (Jiang *et al.*, 2014; Zhang *et al.*, 2015).

QAC can also be implemented during query reformulation. Jiang *et al.* (2014) have analyzed how users reformulate their queries and then proposed a supervised approach to query auto-completion during the user's reformulation process, taking into consideration three levels of reformulation-related features: term, query, and session. Results show a significant improvement in predicting users' queries successfully over

The version of record is available at: http://dx.doi.org/10.1561/1500000073

baselines. Cai and Rijke (2016) have proposed further personalizing query suggestions by adding lexical and semantic information gleaned from the searcher's long-term and current-session query history, but only achieved marginal incremental improvement in query performance. The risk of suggesting queries based on the searcher's query history in the current session during query reformulation is that it will fail if the searcher changes the topic of their search (Cai *et al.*, 2014).

The usage of QAC has been investigated through log analysis. Mitra *et al.* (2014) found through an analysis of the Bing query log in 2004 that QAC usage rates vary considerably with search topics, with popular topics and navigational queries making the most use of it. Hofmann *et al.* (2014) found an average usage rate of 29% for first queries only. Smith *et al.* (2016) reported a 26% rate across all queries within sessions in a user experiment. Further analysis (*e.g.*, Smith *et al.*, 2017) found the real value of QAC was in shorter sessions where there was a noticeably higher retrieval rate. When QAC was used, it was most likely to have been in the first query of a session, but greatly diminished in subsequent queries. If the first query of a session did not use QAC, then it was far less likely to be used in subsequent queries.

3.2.3 Content-Based Multimedia Search

Today, besides text retrieval, there are many different types of search systems that allow users to search for a variety of formats of information, for instance, image, video, computational knowledge, and argumentation, to name a few. Content-based image retrieval systems allow users to sketch coarsely detailed pictures and retrieve similar images based on different features. For example, the RevIMG (http://www.revimg.com/) system allows users to search by texture similarities. TinEye Reverse Image Search (http://www.tineye.com/) allows users to submit an image to find out where it came from, how it is being used, if modified versions of the image exist, or to find higher resolution versions. TinEye regularly crawls the web for new images and also accepts contributions of complete online image collections. Today's search engines even allow users to search for images or videos with an image or video file as a request. For example, VideoQ (Chang *et al.*, 1998) was the first to develop the first

The version of record is available at: http://dx.doi.org/10.1561/1500000073

online video search engine supporting automatic object-based indexing and spatiotemporal queries. IBM researchers proposed the QBIC system that relied on the query of image and video content (Flickner *et al.*, 1995). Another type of search system allows searching by knowledge graphs, such as the Wolfram—Alpha (http://wolframalpha.com). It has a vast collection of built-in data, algorithms, and methods from which it can generate immediate results for either Mathematica input or any freeform linguistic input. This system is believed to make many programming and development tasks much easier. Recently, a new type of search system has appeared, an argumentation search (http://argumentsearch.com). When a user searches a topic with it, the search system finds and summarizes the pros and cons of the topic in real-time.

3.3 The Evaluation Subprocess

The evaluation subprocess includes the evaluation of SERPs, the visualized search results, and the within-document retrieval.

3.3.1 The Evaluation of SERPs

The execution of a search action involves the usability issues of search user interfaces.

Hearst (2009) provides an overview of the presentation of a SERP. Historically a SERP has been presented as a list of search results, each including a title, the page URL, and a summary (or so-called "snippet") of the content of the page. Much research has been conducted concerning the visual representation of these captions (*e.g.*, Aula, 2004; Cutrell and Guan, 2007; Rose *et al.*, 2007).

Without an informative snippet, even the most relevant document may not be clicked or browsed by searchers. Turpin *et al.* (2007) investigated how accounting for the summary judgment stage can alter IR systems evaluation and comparison results. After retrieving 150 search queries issued by middle school children, Bilal and Huang (2019) compared the readability and the level of word complexity of the SERP snippets and that of their associated web pages between Google and Bing. Their research showed that the readability of Google SERP snip-

The version of record is available at: http://dx.doi.org/10.1561/1500000073

pets was at a much higher level than those of Bing. The readability of the snippets generated by both engines mismatched the reading comprehension level of children in grades 6–8. It pointed out the necessity of considering searchers' reading comprehension ability when generating the snippets, especially for young users. On SERP pages, manual experimental results (Iofciu *et al.*, 2009) and eye-tracking studies by (Savenkov *et al.*, 2011) suggest that query term highlighting would help draw the searchers' attention to the results that are most likely to be relevant to the query. Zhang (2018) compared several snippet text highlighting strategies through a user experiment. She carefully designed four highlighting strategies: the original highlighting strategy (query terms were all highlighted in the snippet), the reduced highlighting strategy (only the longest three query words were highlighted), the task-level highlighting strategy (only the terms highlighted by at least 5 users in the task were highlighted) and the result-level highlighting strategy (the terms highlighted by at least 4 users in each snippet result). She conducted a user experiment with 36 participants using a between-subject design and evaluated the search efficacy using the cost-benefit framework. Her results show that the result-level highlighting strategy can reduce search cost significantly in informational search tasks and transactional tasks, but not in navigational tasks. Snippets have been constructed in different ways, either by humans or by automatic construction methods. Bando *et al.* (2010) investigated how 10 humans constructed snippets for four queries related to two documents. The researchers observed that whereas participants extracted the same pieces of text around 73% of the time when creating their extractive queries, the automated approaches only used these same fragments 10% of the time.

3.3.2 Visualization of Search Results

User interfaces designed to support document selection or browsing of search results can be distinguished by textual and visual representations. Most IR systems such as modern search engines present the search results as a list of documents, ranked by order of relevance. There have been some IR systems that display the search results with visualization interfaces. The purpose of the visualization of search results

The version of record is available at: http://dx.doi.org/10.1561/1500000073

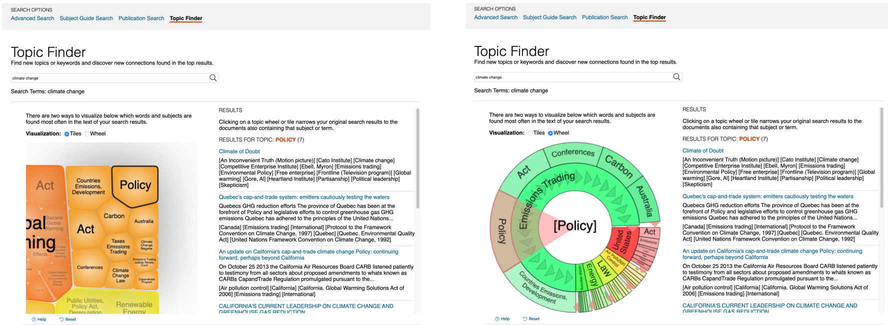
by such engines as FeatureLens (Don *et al.*, 2007), TileBars (Hearst, 1995), and Jigsaw (Stasko *et al.*, 2008) is to provide searchers with the inner structure of a huge text and help searchers mine meaningful and comprehensible context. Ahn and Brusilovsky (2013) proposed and implemented the adaptive visualization search system, Adaptive VIBE, in which they incorporated interactive visualization into personalized search based on the Vibe system. They conducted a user study to evaluate the performance of this system and found that Adaptive VIBE was able to improve the precision and the productivity of the personalized search system because the system could help searchers to discover more diverse sets of information.

In much the same vein, the Topic Finder feature of Gale online databases has the option of displaying titles of retrieved results in either a tile or wheel format (see Figure 3.2). To present the hierarchical relationships of bibliographic records in IR systems more effectively for users, Merčun *et al.* (2017) compared four different hierarchical layouts (indented tree, radial tree, circlepack, and sunburst) for presenting the complex relationships within work families (*i.e.*, the relationships among the entities and their derivatives in describing the work). The findings of user evaluation indicate that the indented tree and sunburst layouts are most successful in terms of search performance and user perception while the hierarchical layouts are useful for work families and exploratory tasks (see Figure 3.3).

3.3.3 Within-Document Retrieval

User interfaces designed to support user access to segments of full-text documents have been discussed in the research literature as part of the book selection process (Wacholder and Liu, 2008; Wacholder *et al.*, 2006), within-document retrieval (Harper *et al.*, 2004), focused retrieval (Arvola *et al.*, 2012) or document triage (Buchanan and Loizides, 2007; Loizides *et al.*, 2014). For example, Harper *et al.* (2004) proposed a user interface called ProfileSkim that provides an interactive bar graph for retrieving relevant segments of a document. The findings from a user study in which users performed manual indexing tasks reveal that the proposed interface is more efficient than the control interface having the 'Find'

The version of record is available at: http://dx.doi.org/10.1561/1500000073



(a) Tiles search results visualization **(b)** Wheel search results visualization

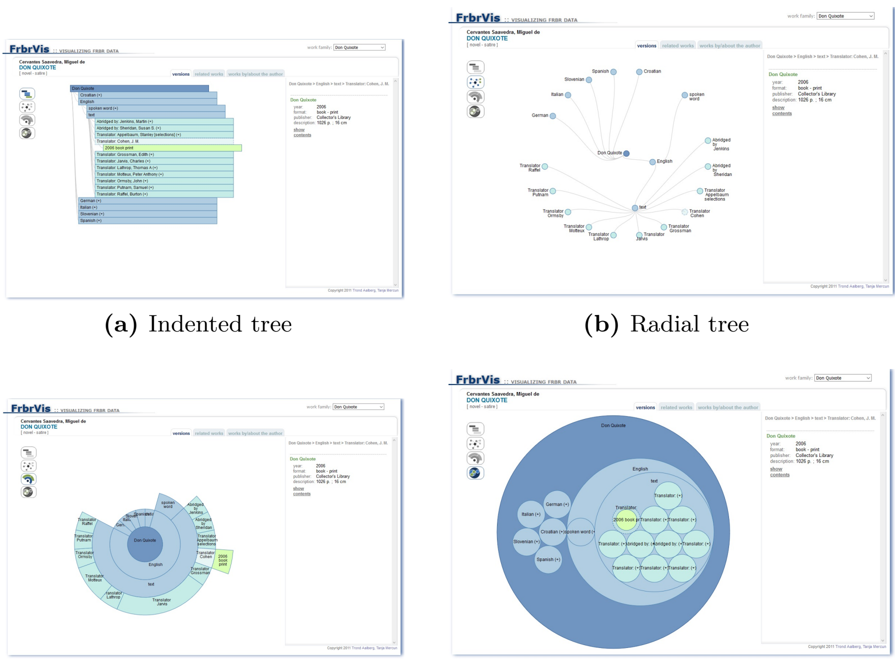
Figure 3.2: Two types of search results visualization interfaces to support document selection as displayed by the Gale Topic Finder.

command in a web browser, and it is at least as effective as the 'Find' command tool. Schwartz *et al.* (2010) proposed a Focus+Context user interface, an extension of TileBars (Hearst, 1995) for within-document search and navigation. A usability study of the proposed interface suggests that there are significant differences in search times for the different visualizations. Loizides *et al.* (2014) indicate that the search tool of matched query term highlights was rarely used in the process of making relevance judgments for documents. Overall, these studies suggest that interfaces with visualizations of term distributions in a long document can efficiently support user access to portions of the document.

Gutwin *et al.* (2017) revealed that user interfaces of spatially stable overviews of the entire document are efficient for document navigation. The results from field experiments reveal that overviews support both pattern matching and revisitation, and the user interface is particularly useful for finding previously visited pages. To build spatial memory of long documents, Mollashahi *et al.* (2018) proposed user interfaces with augmented scrollbars that use visual items as landmarks to support revisitation in long documents. The findings show that double-icons and two-level augmented scrollbars using icons lead to better search performance and score high on user preferences.

From a practical perspective, the user interfaces of PDF viewers,

The version of record is available at: http://dx.doi.org/10.1561/1500000073



(a) Indented tree **(b)** Radial tree

(c) Circlepack **(d)** Sunburst

Figure 3.3: Four different hierarchical layouts for displaying the relationships among bibliographic records (Merčun *et al.*, 2017).

which have implemented the search function for accessing segments of a full-text document, have displayed the matched query results differently. For instance, Figure 3.4 illustrates the Find function (ctrl+F) in Acrobat Reader displaying the query matched results by a simple search box in the top right corner; a full search option displays the matched query results by page order on the left after digging into the Find function. Figure 3.5 is an example of a Preview pane in Mac OS in which the matched query results sort orders are juxtaposed as search rank (the system default option) and page order at the top left. Figure 3.6 shows how the visualization of matched term distribution in the scrollbar reveals the location of the terms in the document and their density.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

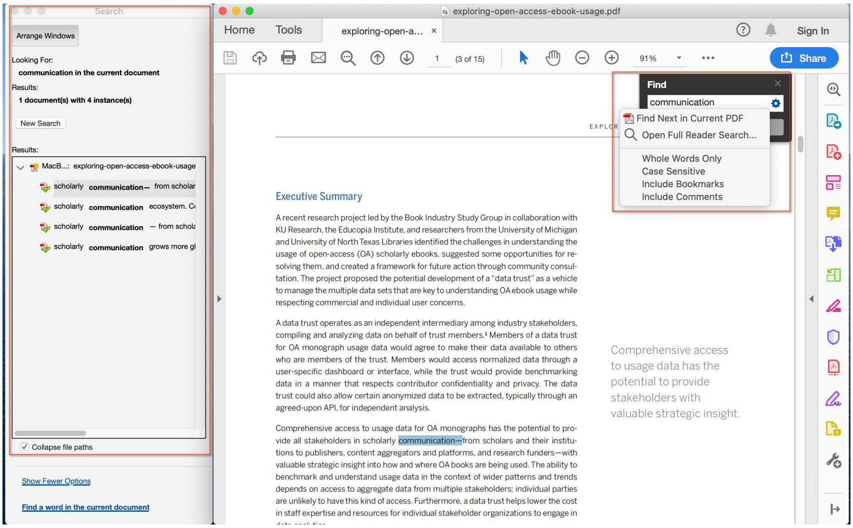


Figure 3.4: Keyword search and navigation functions in Acrobat Reader support within-document retrieval with full reader search.

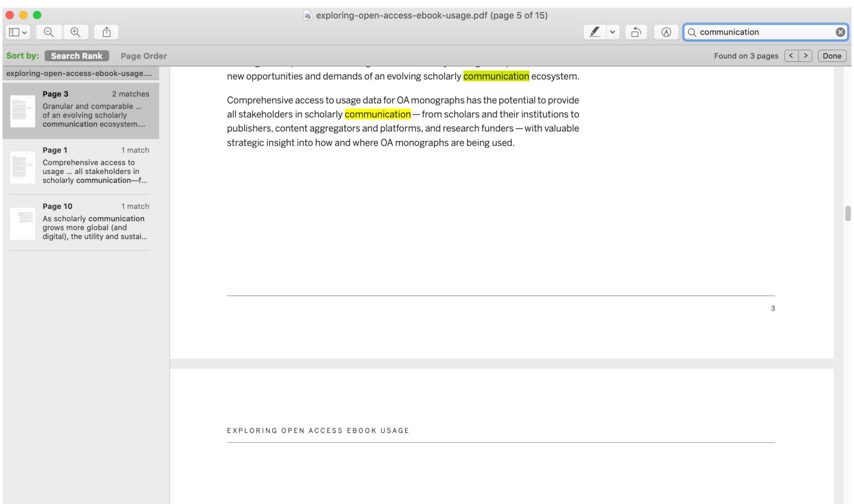


Figure 3.5: Keyword search and navigation functions in Preview mode support within-document retrieval with search results sorted by search rank or page order.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

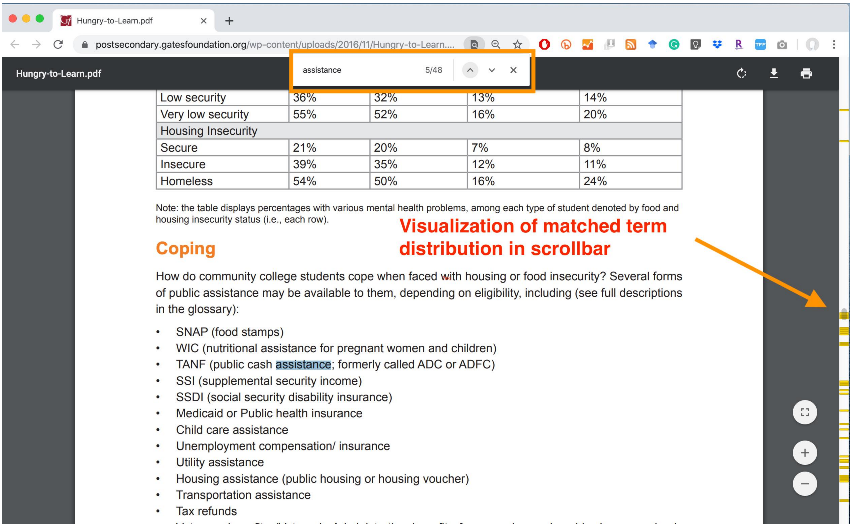


Figure 3.6: Keyword search and navigation functions of PDF.js support within-document retrieval with a visualization of matched term distribution in the scrollbar.

3.4 The Use Subprocess

In regard to use, researchers have explored various aspects including searchers' information use behaviors, the relationship of searchers' information use behaviors with information search behaviors, the discovery of system features that could help in supporting information use and work task performance.

Some effort has been made to make use of search engines in the support of work tasks through searching. One way is to embed search engines in the work task setting. For example, Brandt and colleagues (Brandt *et al.*, 2009; Brandt *et al.*, 2010) dealt with online sources in support of programming. Brandt *et al.* (2009) examined programmers' searching and code writing behaviors and found that they used web search and web sources for several purposes: for just-in-time learning and for gaining high-level conceptual knowledge through web tutorials, as a "translator" for exact terminology or syntax through a web search, and as external memory for complicated syntax that can be accessed as needed, and so forth. Brandt *et al.* (2010) further designed an interface called Blueprint that embedded a task-specific search engine in the code

The version of record is available at: http://dx.doi.org/10.1561/1500000073

development environment which was able to enable programmers to write better code, find example code faster than with a standard Web browser, all of which changed how and when programmers searched the Web.

Another way that search engines are used is to support low level language-related tasks, such as spell checking, grammar checking, disambiguating between homonyms, defining, and providing pronunciations. Fourney *et al.* (2017) found that language-related searches were indeed common, accounting for at least 2.7% of the queries in their data set. They also found that the convenience of search engines, the richness of the information search engines can provide, and the familiarity and confidence users have with their search engine use contributed to the common use of search engines as language tools.

3.4.1 Searching for Writing, and Other Information Use Behaviors

In a study by Kumpulainen and Jarvelin (2010) examining work-task-driven information access in the molecular medicine domain, multiple sources and channels were found to be used in information seeking including web search engines, literature databases, websites, and bio databases. This led to the suggestion of setting information integration as the ultimate goal of system development, with an interim or parallel strategy of exploring how single information systems can assist information use.

Whereas searching behaviors in the context of having or not having a work task have been explored quite extensively, those in the context of generating a task product, such as a written text, have been explored very little. Along the latter line, Liu and Belkin (2012) found that the ratio of pages useful to all, but not the number of queries or web pages found, positively correlated with task performance. They also found that although time spent searching positively correlated with the number of queries and web pages found, it did not correlate with work task performance. Hagen *et al.* (2016) found that in non-fiction text writing, users would often submit recurring anchor queries to avoid losing the main themes or to explore new directions. Fact-checking queries were found to often conclude a writing task, but the number of submitted queries was not a good indicator of task completion.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Furthermore, information use is so different from information search that their behaviors and performance do not even correlate. Vakkari and Huuskonen (2012) found that effort expended in the search process degraded search precision but improved the essay writing task outcome. Effort in expanding Medical Subject Heading (MeSH) terms within search sessions and effort in assessing and exploring documents in the result list between the sessions degraded precision, but led to better task outcome. Thus, human effort compensated bad retrieval results on route to good task outcomes. Liu and Belkin (2012) found that the performance of report writing did not correlate with those measures typically used in evaluating search performance, such as the number of useful webpages viewed, the number of queries issued, or the length of time required to complete the task, but did correlate with users' knowledge of task topic, their previous experience on the task type, their effectiveness in finding useful web pages, and the amount of time they could allocate to writing. Liu and Belkin (2014) further found that users with different levels of task topic knowledge performed differently on different tasks. Particularly, users with higher-level knowledge tended to perform better in the parallel-structured task (subtasks being in parallel with each other) than in the dependent-structured task (some subtasks being dependent upon the completion of others); in comparison, users with lower-level knowledge tended to perform better in the dependent-structured task than in the parallel-structured task.

More recently, Vakkari *et al.* (2018) attempted to predict the success of IR from information use behaviors for writing tasks, together with search behaviors such as querying and clicking. Two main indicators of the usefulness of search results were: the number of words reused from the clicked search results, and the number of pastes. Increased search result usefulness was also found to correlate with a decrease in effort spent on editing the pastes for the essay, which was consistent with findings in Vakkari and Huuskonen (2012). These results all indicate that in order to design systems that support work task accomplishment, information use should be further explored and better understood.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

3.4.2 In Support of Information Saving and Collecting

Liu *et al.* (2018) explored which system features it is that college students use to find and save information when they work on their real-life tasks. They found that some generic features among systems are helpful for information saving, for example, downloading or emailing information to oneself, while, other features are specific to only one or to a class of systems. For example, whereas library systems often let searchers save resources in a personal list that can be e-mailed or saved to an account, Pinterest lets users pin items to their boards.

Commercial search engines have attempted to provide more functions to help with user tasks than simply returning search results. *Google – Notebook FAQ* (n.d.), Yahoo! Search Pad (Needleman, 2009), and Microsoft Thumbtack (Brown, 2008) are examples of approaches to enabling the system to automatically save search results or notes rather than requiring the user to manually email them to himself. Unfortunately, these functions have been shut down due to not attracting much use. Regardless, except for the deceased Google Notebook, which allowed users to take their own notes, all other services are basically about information saving and collecting rather than information use.

To the best of our knowledge, current commercial search systems do not generally show features that help searchers use the information to create task outcomes and accomplish their tasks, and any that attempt it, do not do so extensively. One reasonable explanation is that system features supporting task accomplishment may depend, to a large degree, on the specific tasks attempted and their specific task requirements. For example, writing an essay differs from finding a restaurant or creating a vacation plan in many aspects, such as the type of information sources consulted, the format of the final products, and the writing/creating process. The specificity poses a challenge to generic search engines.

3.4.3 In Support of Note Taking

Some approaches have attempted to assist users' work tasks by being better at providing search results or documents. Budzik and Hammond (2000), for example, designed a system named Watson that could suggest documents as a searcher writes a paper. Golovchinsky *et al.* (1999) used

The version of record is available at: http://dx.doi.org/10.1561/1500000073

readers' annotations on documents as queries. However, these are still along the direction of supporting information search.

Among the few approaches to supporting information used, some work by aiming at helping users take notes. In proposing interface approaches in support of semantic navigation, Kopak *et al.* (2010) suggested one means, that is, providing a closer coupling between the reading and writing process by fostering annotation. This is indicative of an interface design that supports writing as the work task outcome. Along with reading and comprehension, the making of annotations should also be fostered as this could also be helpful to information use since the annotation pieces could be used in the generation of task outcomes such as article or report writing.

System interfaces that support annotations or note taking as part of the exploratory search process have been designed and evaluated. Ahn *et al.* (2008) designed an interface called TaskSieve that supports task accomplishment. Specifically, it provides a "Task Model Notes" panel on the right side of the search interface into which users can paste and type notes. Similarly, He *et al.* (2008) designed an interface that provides a shoebox in which users can assemble text fragments. Although there have not been any approaches or studies aimed at helping with information use or note taking in the past decade, in recent Search as Learning studies, experimenters have often designed search tasks that involved a learning output (*e.g.*, writing an essay or answering questions) from which they could evaluate learning performance after search (Wilson and Wilson, 2013; Liu *et al.*, 2020a). In addition, it is important to ensure that search systems support not only note taking, but the whole task completion process (White, 2016). A recent study by Roy *et al.* (2021), examined the effect of such notepad-like interfaces on users' knowledge and understanding during a complex, learning-oriented search task. It was found that participants using an interface with highlighting enabled covered 34% more subtopics, and participants that used a search interface with note-taking functions enabled covered 34% more facts in their essays when compared to those using a standard web search interface. This study also found that incorporating active learning tools significantly changed the search behaviour of participants across a number of measures. More investigative research is needed into

The version of record is available at: http://dx.doi.org/10.1561/1500000073

how to implement highlighting and note-taking functions in support of users' search processes, especially in learning-oriented contexts.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

4

Personalization and Contextualization of Search Interfaces

Personalizing and contextualizing search interfaces adapt the information to personal and/or situational needs for an improved information experience (*e.g.*, a filtered or re-ranked search result list, interest-related shopping recommendations, and selected or summarized facts about a historical event). Increasing information quality, for example by effectively querying, filtering, ranking, presenting, or otherwise processing information based on peoples' dispositions, abilities, and traits (*e.g.*, age, gender, user group membership, (dis)abilities, and interest) and/or their environmental (or external) situation (*e.g.*, location, time, domain-related tasks, and device information) can be helpful in achieving this goal. While context models may contain all types of information, personalization generally adapts information experiences based on personal attributes while contextualization uses situational information.

This section reviews research on adaptive search interfaces that fit in one (or both) of these categories. We review work in the area of user modeling, adaptation and personalization, human-centered computing, HCI, and IIR. We specifically categorize the related work into two main areas, those of personalization and contextualization. In addition, we also discuss how search interfaces have been designed specifically for

The version of record is available at: http://dx.doi.org/10.1561/1500000073

children, for older adults, and for people with disabilities to reflect that there are certain groups of people that need a customized search interface.

4.1 Adaptive User Interfaces

The objective of the adaptive user interface is to provide users with interactive support that can adapt to their tasks, performance, intentions, goals, and characteristics. The sources of signals used for inferring the user states include user search behavior, eye gaze and other physiological signals, the device platform on which they are currently working, and the user's contextual environment. Personal signals (*e.g.*, the estimated expert level of a user regarding a topic) and the user's behaviors (*e.g.*, the implicit feedback on the relevance of search results) are often referred to as personalization or personalized adaptation. External signals outside the user's control are often referred to as contextualization or context-aware adaptation (*e.g.*, using one's current location when searching for shops to filter out results that are outside of normal driving distance).

With a focus on IR system design, these signals have been used to infer document relevance, user information needs, user intents, and user characteristics. Based on the modeling using features derived from the signals, implications and recommendations for user interface design have been suggested in support of specific tasks. From a system design perspective, research on gaze-based search have focused on the use of eye gaze data to infer user characteristics and search tasks. These observations correspond to the application domains of eye trackers: interactive and predictive, where the interactive domain is considered a replacement for the mouse and the predictive has more sophisticated user modeling techniques (Liu and Bierig, 2014). As revealed in the following review, studies of adaptive user interfaces appear in several fields of study, including IIR, information visualization, HCI, user modeling and personalization, and human-centered computing.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

4.2 Personalization of Search Interfaces

Related work on personalization is characterized by search behavior, personal attributes, and user perception.

4.2.1 Personalization by Search Behaviors

In the context of adaptive user interfaces, researchers have used explicit relevance judgments to build user interest models and presented the search results in a manner showing keyword-to-document relatedness in order to facilitate document understanding (Ahn and Brusilovsky, 2013). Adaptive Vibe (Ahn and Brusilovsky, 2013), for example, provides a personalized and exploratory search interface that filters and visualizes hundreds of results based on an evolving user profile about their proximity to topical themes by the documents selected and user annotations. Medlar *et al.* (2017) proposed optimizing the exploration rate based on the user's search behaviors such as clicks and reading time, and personalizing document ranking based on the user's familiarity with the topics. Using measures of search behavior to separate the exploratory from look-up search tasks, Athukorala *et al.* (2016) found that the most distinctive indicators are query length, maximum scroll depth, and task completion time. These studies all used search behavior to infer user models, but search behavior alone may not be able to fully consider the complexity of user mental models.

Another thread of research is concerned with the use of eye gaze data to predict the user's intentions, goals, and tasks, with a bit more sophisticated user mental models. Similar to the search actions, visual search behavior is also a reflection of the user's cognitive states. For instance, to develop real-time interactive systems, Karaman and Sezgin (2018) used eye gaze to predict user task-related intentions and goals in which manipulation tasks in the user interfaces, such as scrolling, free-form drawing, resizing and dragging, are supported. Low *et al.* (2017) used eye gaze data as implicit feedback to support the user's intentions and goals during the search process for exploratory search tasks. It was found that the eye gaze metrics of number of fixations and fixation durations as well as pupil dilations are good indicators of

The version of record is available at: http://dx.doi.org/10.1561/1500000073

whether people are looking at the target image or not. Steichen *et al.* (2014) used eye gaze data to infer visualization task types, such as task complexity and difficulty. The findings suggest that AOI (area of interest) related features are important for producing more accurate predictions of visualization type, user performance, and user cognitive measures. Spiller *et al.* (2021) developed computational models to predict visual search tasks by using the feature set of fixations, saccades and pupil diameter and deep learning models for time series classification. The findings suggest that it is feasible to infer the users' task success within the first ten seconds of interaction. Overall, these studies reveal that eye gaze metrics and pupil size are good predictors of user intentions, goals, and tasks, particularly when using deep learning models. However, support for the connections between user mental models and user search performance need to be clarified.

In addition to the search actions and eye gaze data, recent studies have focused on the use of multiple physiological signals, such as electroencephalography (EEG) and eye gaze to model user intents. For example, Ruotsalo *et al.* (2018) developed an interactive intent modeling system for aspectual exploratory search tasks. The main feature of this intent model is the visualization the user's search intents as keywords in combination with interactive visualization. Further development of intent modeling has used the neurophysiological signals of EEG and eye movements as relevance feedback (Jacucci *et al.*, 2019). To develop user-adaptive systems that can support user's intents, intentions, and goals during the search process, researchers have used the primary task context (*e.g.*, writing task), user interaction data, and physiological signals to model the user's search intents. Koskela *et al.* (2018) used implicit primary task context in writing tasks to model user's search intent by providing a list of suggested keywords based on the intent model. These studies indicate the use of multi-modal features from various sources such as eye gaze, user interaction, and physiological signals can be useful for user intent modeling, but how user search intent modeling contributes to user search performance in different task contexts needs further research.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

4.2.2 Personalization by Personal Attributes

In the context of search interface design, personal attributes refer to user characteristics, such as personality traits, cognitive abilities, cognitive styles, and individual differences that can affect user interactions with information visualization systems and other search interfaces. From an IIR perspective, research shows that search experience is correlated with the recall measure (Liu, 2010) and experienced users with a high level of perceptual speed achieve better search performance (Al-Maskari and Sanderson, 2011). In an extensive review of adaptive and personalized visualization studies, Ottley (2020) proposed that personality traits and cognitive abilities can serve to modulate intent modeling based on user interactions with information visualization systems, and saw the inference of traits from user interactions as a promising research area.

Steichen *et al.* (2014), for instance, used eye gaze data to infer cognitive abilities, such as perceptual speed, visual working memory, and verbal working memory. Their results show that classification accuracy using a relatively simple set of area of interest (AOI) features is higher for visualization task types than that for cognitive abilities. Research on the relationship between user characteristics and eye-tracking measures has suggested that user's cognitive abilities, such as perceptual speed and verbal working memory, are correlated with eye-tracking measures (Toker *et al.*, 2017). In a study that assesses the impact of English reading comprehension ability on textual documents with embedded visualizations (Toker *et al.*, 2019), it was found that people with lower reading abilities need more transitions from text to visualization and take significantly longer to fixate on the relevant bars within the visualization. These studies suggest that cognitive abilities can be inferred from user interactions with information visualization systems, particularly from the eye gaze data. However, how adaptive information visualization systems can support people having different levels of cognitive ability deserves further research.

In addition to cognitive abilities, the user characteristic of cognitive styles are correlated with eye-tracking metrics. For example, in a study of visual search activity with tasks of varying task complexity, Raptis *et al.* (2017) showed that field-dependent (holistic) users produce more

The version of record is available at: http://dx.doi.org/10.1561/1500000073

eye gaze transitions than field-independent (analytic) users. When task complexity increases, the attention distribution of field-dependent users by AOI is more equally distributed than that of field-independent users. The results of a study using eye-tracking metrics to predict the user's cognitive style for graphical password composition tasks (Katsini *et al.*, 2018) show that the saccade length is the most effective eye-tracking metric for predicting field-dependent (*i.e.*, holistic) users whereas the fixation count is the most effective for field-independent (*i.e.*, analytic) users. Overall, these studies reveal that there are significant correlations between eye-tracking metrics and cognitive styles. Further research needs to consider how system-driven customizations can be provided to better support the specific search tasks.

Research on personalization by personal attributes has also been concerned with the relationship between user characteristics and features of search interfaces. For instance, in an eye-tracking study of different search interfaces for domain experts (Liu *et al.*, 2017), it was found that there are significant interaction effects between types of search interfaces and cognitive styles by the proportion of fixations in reading time for a specific interface component. Analytic users were more attracted to a simple search interface similar to the Google search engine whereas holistic users preferred a more complex search interface similar to some widely used online search databases for academic articles, such as EBSCOhost. Drawing from information foraging theory, it was found that people with different cognitive styles use different search strategies as observed by eye gaze behavior in terms of AOI-based fixation time in an uncertain environment perceived as difficult (Wittek *et al.*, 2016). The user's experience with the search system affects his eye gaze patterns in terms of AOI-based fixation time within the various interface components (Liu *et al.*, 2017; Tang and Song, 2018). These studies suggest that eye gaze data can be used to predict the user's cognitive style and search experience as part of a user mental model for developing user-adaptive or natural search interfaces. Further research needs to consider building computational models like deep learning models (Spiller *et al.*, 2021) that can adapt to users with varying cognitive styles and search experience for enhanced search performance.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Another thread of personalization by personal attribute research has focused on system-driven customization of user interfaces. For example, Lallé and Conati (2019) demonstrated that users benefit from a system-driven customization of the information content presented in an information visualization system dependant upon the user characteristics of visualization literacy and locus of control. Christmann *et al.* (2010) proposed a gaze control search interface in which distortions of the images are adapted to the user's visual capabilities. To design systems that adapt to the user's mental effort and user performance, Buettner *et al.* (2018) found that difficult search tasks contribute to more pupil diameter variability, which is conceptualized as a measurement of interest. Moshfeghi *et al.* (2019) found significant correlations between the detection of user information needs and the brain signals from fMRI in the context of proactive search engines. These studies suggest that system-driven customization can provide proactive assistance before users recognize their information needs and interests. However, the issue of the balance between user control and system customization for search task completion needs further study.

4.2.3 Personalization by User Perception

From an IIR perspective, user perception is concerned with the user's interpretation of the usability of search interfaces through his sensory system. For example, researchers have investigated the use of physiological signals as implicit feedback to detect the user's perceived relevance of documents and to infer user information needs. Using the signals from eye movement and EEG, Gwizdka *et al.* (2017) demonstrated that user-perceived relevance of documents can be inferred from these signals. Oliveira *et al.* (2009) identified the perceived relevance of web search results by pupil size and noted the challenges of noise using pupillometry. Barral *et al.* (2015) showed that, with the correct selection of features and time windows, signals from electrodermal activity (EDA) and corrugator supercilii activity (CSA) can predict the perceived relevance of documents. González-Ibáñez *et al.* (2019) used features from multiple modalities for the detection of perceived relevance in IIR. It was found that the feature of the left mouse click contributes the most

The version of record is available at: http://dx.doi.org/10.1561/1500000073

to the distinction between relevant and non-relevant web pages. Barral *et al.* (2016) explored the relationship between the physiological signals and the perceived relevance of the text read, as well as the affective responses after reading the text. Eugster *et al.* (2014) found that the relevance of terms that represent a pre-determined search topic correlates with brain activity as measured by EEG, even without the use of other user interaction data. Overall, these studies suggest that physiological signals can be used to detect the perceived relevance of documents. However, more research is needed for the implementation and evaluation of how the implicit feedback provided by systems can support users in task completion.

As revealed in Oliveira *et al.* (2009), it is challenging to use pupillometry as an indicator of cognitive activities. Marshall (2002) proposed the Index of Cognitive Activity (ICA) as a measure of cognitive workload based on pupil dilation. Ehlers *et al.* (2018) showed that explicit pupil size changes can be used as a selection mechanism for user interface design. Pauchet *et al.* (2018) proposed adapted interaction modality based on gaze direction for the design of the touch surface. In a study that was designed to identify the user characteristics that significantly affect people's processing of textual documents with embedded visualizations, Toker *et al.* (2018) found that the user characteristics of need for cognition significantly affect users' perceived ease-of-understanding and interest, and the user characteristic of spatial memory has a positive correlation with ease-of-understanding. Overall, eye gaze data, such as the changes in pupil size and gaze direction, together with user perception and performance data can be useful for developing adaptive user interfaces.

Another thread of research has focused on the relationship between the features of interaction data and people's emotional states. For example, Edwards and Kelly (2017) demonstrated that user interaction data together with physiological responses like heart rate and galvanic skin response can distinguish between engagement and frustration in search sessions. Kiseleva *et al.* (2016) used features from the use of mobile devices and interaction with intelligent assistants to model user satisfaction. Wu *et al.* (2019) analyzed eye movement data and made connections between eye gaze patterns and user satisfaction. Sarraf (2019)

The version of record is available at: http://dx.doi.org/10.1561/1500000073

identified the brain areas that correlate with the affective dimensions of Kuhlthau's ISP model using the EEG. Sarraf (2019) identified the brain areas that correlate with the affective dimensions of Kuhlthau's ISP model using the EEG. These findings suggest that user interaction data together with eye gaze data or physiological signals can be used to infer the user's emotional states, such as user satisfaction and frustration.

4.3 Contextualization of Search Interfaces

Related work on contextualization is characterized by search behavior and situation.

4.3.1 Contextualization by Search Behavior

Research on contextualization of search interfaces by search behavior is characterized by examining the relationship between the behavioral signals and the search contexts. From an IIR perspective, research suggests that search tasks and contexts affect how users engage with the search processes when interacting with IR systems (*e.g.*, Li and Belkin, 2008; Järvelin *et al.*, 2015; Tamine and Daoud, 2018). The majority of IIR studies in the 2010s have focused on search tasks that instruct participants to search for information for a writing task; the predictors of the usefulness of search results were search behavior, characteristics of tasks, and search interface (Vakkari, 2020).

From a system design perspective, research has focused on the design of adaptive user interfaces for different contexts. For example, using a mixed-method approach, Wu and Liang (2018) examined the contexts of mobile search and their relationship with mobile application usage. Yigitbas *et al.* (2019) used context monitoring and user feedback to trigger user interface adaptation features adaptive to the user, platform, and environment. These studies indicate that the contexts of mobile use (platform, user, and environment) and associated features can be useful for triggering user interface adaptations, which affect user-perceived performance, user preferences, and user satisfaction. However, further research on the relationship between search tasks and contexts of mobile use is needed for developing adaptive user interfaces.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

4.3.2 Contextualization by Situation

Research on contextualization by the situation is characterized by examining the relationships among the situational context, user search behavior, and user preference. In this type of research, the context of the user's physical environment during mobile phone use are correlated with search tasks. For instance, Aliannejadi *et al.* (2019a) used a mobile application to capture the situational context when users perform a search. It was found that the level of user engagement decreases when users are in the walking context or on transport. Al-Ismail *et al.* (2019) found a significant relationship between user preferences and contexts (*i.e.*, physical and social) in mobile learning. Harvey and Pointon (2017) showed that fragmented attention induced by common mobile situations, such as walking, significantly affects people's perceptions of performance, but in actuality, there is no significant difference in objective performance.

Unlike the majority of IIR studies in the 2010s that focused on search tasks instructing participants to search for information for a writing task (Vakkari, 2020), research on contextualization by situation has been concerned with how the situational factors affect user search behavior and performance. For example, one study of search situations like the writing task revealed that querying behavior, mouse-clicking, and text editing can predict the retrieval success by building linear regression models in which the primary indicators were the number of words reused from the clicked search results and the number of pastes (Vakkari *et al.*, 2018). Salminen *et al.* (2020) demonstrated that the work roles of professionals (marketing professionals and data analysts) correlate with their visual search behavior when interacting with AI-driven persona interfaces. Specifically, the use of numbers in the search interface increased the perceived usefulness for data analysts. Overall, these studies extend our understanding of how situational contexts, such as writing tasks and work roles, affect user search and visual behavior for task performance.

Another research area that fits within contextualization by situation is collaborative information seeking and retrieval, which is defined as "the study of the systems and practices that enable individuals

The version of record is available at: http://dx.doi.org/10.1561/1500000073

to collaborate during the seeking, searching, and retrieval of information" (Foster, 2007, p. 330). For example, through a study of an intensive information domain, Hansen and Jarvelin (2005) propose a refined IR framework that includes a collaborative aspect in information seeking and retrieval processes by task stages. González-Ibáñez (2015) propose the use of technologies, such as electrodermal (EDA) activity and eye-tracking, to study the affective dimensions of collaborative information seeking. Elbeshausen *et al.* (2015) compare work-based and leisure collaborative information seeking by identifying different phases of collaboration. Given and Willson (2015) propose the concept of parallel work, individualized framing of collaborative information behaviors through an investigation of digital humanists in research contexts. Fidel (2012b) advocates a cognitive work analysis project to examine the relationship between the human cognitive processes and the complexities of collaborative work environments. Dörk *et al.* (2021) adapt a co-design framework involving actors, activities, and artifacts to the design of information visualization systems able to collaborate with a specific domain group. These conceptual frameworks have guided collaborative information seeking and retrieval research by considering the collaborative aspects of human information seeking behavior and interaction for the practice of system design as well as for understanding collaborative information seeking in different situations. However, as reviewed below, there is still a gap between these frameworks and the collaborative IR studies.

Research on collaborative IR has focused on how to improve the search results by searcher roles. For instance, researchers have proposed specific techniques to support the concepts of the division of labor and the sharing of knowledge in synchronous collaborative IR (Foley and Smeaton, 2010). An algorithmic mediation approach has been adopted to support small groups of people with a shared information need (Shah *et al.*, 2010). Avula *et al.* (2019) have studied the uses of searchbots in collaborative information-seeking tasks and their effect on search outcomes. It was found that participants can benefit from collaborative awareness, but they are distracted by searchbots during their individual work. More recently, Htun *et al.* (2018) have developed search interfaces to support asynchronous collaborative tasks with unequal access to

The version of record is available at: http://dx.doi.org/10.1561/1500000073

information, with an emphasis on the effect of query and search result awareness on search performance. These studies, however, could be enhanced by considering user search behavior in more realistic settings during the interface design processes as well as the system design frameworks.

4.4 Search Interfaces for Special Populations

In this section, we review related work on search interfaces designed for children, for older adults, and for people with disabilities.

4.4.1 Search Interfaces for Children

Children's access to online resources and services is quickly increasing. In 2020 in the US, 94% of children under the age of 8 years had online access (Common Sense Media, 2020)[1]. The UNICEF report *Growing Up in a Connected World* summarizes data from children aged 9-17 from 11 countries and states that 20-40% of children regularly engage in information seeking to learn and to access news and local information (Winther *et al.*, 2019).

This section takes a closer look at the kind of skills that designers of search interfaces should consider, based on studies synergized by recent review papers and books on the subject. We then present examples of search interfaces that have been created for kids and their special needs.

Gossen and Nürnberger (2013) review relevant aspects of IR that are important when designing search systems for children[2]. In Bruckman *et al.* (2012), the abilities and needs of children are compared to those of adults while the review by Hourcade (2007) provides interface design guidelines for children – both from the perspective of HCI. Below, we summarise findings from these reviews that are most relevant for search interfaces and organise them into *perceptual and cognitive*, *motor*, and *emotional* aspects.

All three aspects should be seen in relation to the individual development of a child. Younger children will generally tend to have more

[1]This is for high-income families. Low-income families in the US have access in 74% of cases.

[2]These findings are reiterated in Gossen (2016, Ch.3).

The version of record is available at: http://dx.doi.org/10.1561/1500000073

issues with cognitively complex operations that require sophisticated motor movements and that are emotionally demanding. Piaget and Inhelder (1969) distinguish four sequential stages of cognitive development: The *sensory-motor* stage represents toddlers of up to 2 years who can interact with objects using only their basic senses. Children in the *pre-operational* stage (2-7 years) start learning a language but have limited attention, are generally self-centered, and possess limited ability in logical thinking but are able to perform simple one-dimensional classification tasks. At the *concrete operational* age (7-11 years), children use trial-and-error problem solving without much ability to work with abstract concepts and hypotheses. From 11 years on, children enter the *formal operational* stage where their abilities match those of adults. Other related theories covering development, perception, intelligence, memory, and such like are discussed in detail in Hourcade (2007).

Perceptual and cognitive aspects: Children in their early sensorimotor and pre-operational stages should not be presented with a text-based search interface due to their lack of reading ability. This age group prefers online games and videos over standard search interfaces. Later, when entering the concrete operational stage, kids start becoming interested in search interfaces. However, they often have difficulties formulating information needs into queries as they lack abstract thinking, language skills, and writing/typing abilities. Their queries have many mistakes, tend to be shorter, and are formed in natural language oblivious to Boolean logic. Often, children have difficulties in typing and need to pay considerable attention to the keyboard. Therefore, support in query formulation helps children to complete and refine their queries, such as by offering spelling correction or query completion features. Real-world categories help children to overcome limitations in abstract thinking. Providing images and voice menus instead of text helps children overcome reading limitations. Simpler texts can provide additional support by making textual descriptions more accessible. Children perform better with interfaces for browsing than for searching as browsing requires less recall knowledge and demands lower cognitive load. However, browsing categories

The version of record is available at: http://dx.doi.org/10.1561/1500000073

should match their age group and interface metaphors should match their knowledge rather than those of working adults.[3] Furthermore, search interfaces should be graphical and when using text, made more readable by the use of large fonts. Large icons, buttons, and other user interface elements in a consistent layout lower the cognitive load and allow children to trace past actions. They should be easily differentiated from their background and indicate interaction. Search interfaces should offer tools for search history and result storage as children may forget past search activity. Menus can be hard to remember and comprehend by children – especially those 7 years and younger – if menu choices are not always visible.

Motor aspects: Children have a reduced capacity for fine-motor control, in addition to smaller hands that make it generally more difficult to use standard mice. Kids often have more issues with keyboards in comparison to mice, but even when using pointing devices, interfaces for browsing are preferable as they can be used through simple interactions. Children have difficulties performing specific interactions with mice such as double-clicking, dragging-and-dropping, scrolling, multi-item selecting (*e.g.*, marquee selection) which suggests that interactions with search interfaces should be kept simple and direct. Additionally, it has been found helpful when (search) interfaces provide large icons and user interface elements (*e.g.*, buttons, or result labels) to support children's reduced precision and dexterity when it comes to motor skills.

Emotional aspects: Children further require search interfaces that provide support for their more demanding emotional spectrum. They require interfaces that are colorful and that integrate multiple forms of media (*e.g.*, sounds, animations, and videos in combination). Children want an interface to be funny and entertaining as well as helpful. They require stronger emotional support to motivate children to be successful with regards to their goals.

[3] The 'folder' is a typical example of a conceptual model that strongly relates with the mind of a working adult: A paper folder in a physical cabinet in an office.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

> Search interfaces should offer strong features for help, such as when a search returns no results. Guidance figures are suggested to provide emotional support and hints. Search interfaces should equally attend to the utility and the entertainment of children. Waiting for a response or a result becomes easily frustrating, so a search interface benefits from offering progressive feedback on ongoing operations (*e.g.*, by animations).

The doctoral thesis, and now a published book, by Gossen (2016) carefully analyzes the specific needs of children for information search. The *Knowledge Journey* is a search engine (see Figure 4.1) with interface elements that specifically support the needs of children aged 7-11. A customizable guidance figure supports children during search and limits frustration. A pie menu, styled as a steering wheel, supports categorical directions for the formulation of search queries. Search results are presented on a sequence of papyrus rolls that offer large clickable areas for selecting results that can then be stored in a treasure chest for additional memory support (Gossen, 2016, Ch.7). In a user study comparing The Knowledge Journey with a classic search interface, participants preferred the colorful and picture-rich search engine over the minimalist design of the Google-style alternative. The mainstream search interface design caused participants to struggle with backtracking, opening results in a separate tab, and locating relevant information in a result list. Most children also had issues with keyboards. A follow-up study with voice input further showed that children preferred a touch interface over voice control. The treasure map interface metaphor of The Knowledge Journey additionally helps to spatially visualize search histories (see Figure 4.1b).

Chao and Lin (2015) studied pre-schoolers' and second-graders' search behavior with a visual search interface. Children intuitively assembled visual queries to search for storybooks by dragging and dropping icons for characters, scenes, and the color of the book onto an empty query book cover (Figure 4.2). The work addresses some of the difficulties that children experience when constructing queries. The visual query interface supports children in their reduced language capabilities and their reduced ability to apply categorical abstract thinking.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

(a) Main search user interface of The Knowledge Journey (from Gossen (2016))

(b) Treasure maps visualise search results (from Gossen *et al.* (2015))

Figure 4.1: Two screens from The Knowledge Journey – a search interface for children

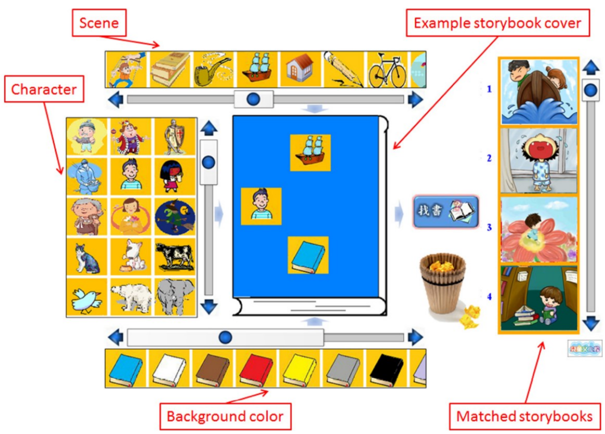


Figure 4.2: Composing a visual search for a storybook based on scenes, characters, and book cover colours (from Chao and Lin (2015)).

Downs *et al.* (2020) investigated spellcheckers as part of search interfaces. As children often tend to select the first option available rather than the most relevant, their work specifically focused on providing child-friendly alternatives. Figure 4.3 shows an example of how their interface uses pictures to help children identify a relevant spelling of a query term.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

While these examples clearly show the benefits of child-friendly search interfaces, there is much to be done. Search interfaces should not only be created with the child in mind, but also their active collaboration (Hourcade, 2007). Fails *et al.* (2013) highlights specific techniques and methods that are suitable for including children in the design process. The annual ACM Interaction Design and Children (IDC) conference[4] brings experts together in developing and researching child-centered designs, interactions, and interfaces that are also highly informative for building better search interfaces.

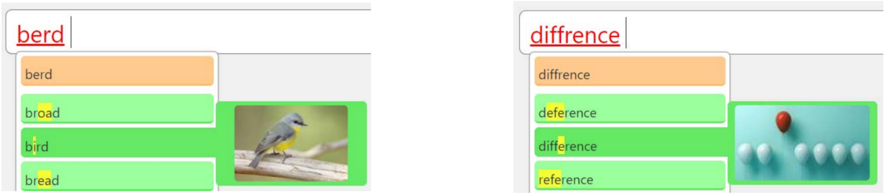


Figure 4.3: Using pictures to strengthen the spelling suggestions for a concrete word (left) and an abstract word (right) (from Downs *et al.* (2020)).

4.4.2 Search Interfaces for Older Adults

Average life expectancy has increased by more than six years between 2000 and 2019 (from 68.8 to 73.4 years) with an increase of 5.4 healthy years (from 58.3 to 63.7) (WHO, 2019). Although the global pandemic has lowered life expectancy (WHO, 2021), the positive trend for longer and healthier lives is expected to continue. In the context of this section, an *aging population* means that more older adults will access search services and their interfaces. Designing for their needs is therefore more relevant than ever before. This section summarizes the most important aspects to consider when designing search interfaces for older adults, and further presents potential alternative directions for search interface design for this user group.

Following the review of 30 studies and articles, Dodd *et al.* (2017) found that older people are mainly challenged in three categories: cognitive abilities, physical abilities, and computer literacy. Cognitive abilities

[4] https://idc.acm.org

The version of record is available at: http://dx.doi.org/10.1561/1500000073

diminish in the areas of attention and memory. Likewise, physical abilities deteriorate over time and often impair senses like eyesight and hearing, as well as motor skills. Furthermore, older adults may be less experienced with computers and less patient with technology if it is not well suited to their purpose or task. Johnson and Finn (2017) describe a wide range of age-related aspects for user interfaces in the first 9 chapters of their book, covering similar aspects at greater length. Similarly, Czaja *et al.* (2019) organize age-related changes in perceptual, cognitive, and movement-related categories from the perspective of human factors engineering while offering recommendations for interface design (see Czaja *et al.* (2019, Chapter 7)). Below, we briefly summarize requirements from these sources by differentiating them into perceptual and cognitive aspects, motor aspects, as well as technological and attitudinal barriers.

Perceptual and cognitive aspects: With increasing age, users experience reduced performance in visual, auditory, and haptic abilities. Cognitive processes like information processing speed, attention, memory, and spatial processing are diminished. Search interfaces should use familiar interface layouts and interaction paths to reduce demands on the user such as forcing them to learn a novel interface. Reduced attention increases the time people need to filter and process the presented information on interfaces. An interface should also address the reduced working memory by not requiring people to remember more than three actions in a task chain toward accomplishing a task, and additionally being reminded to complete future tasks (as they might otherwise be forgotten). Wagner *et al.* (2014) further shows that age may lead to reduced spatial and mental models that need to be considered in search interface design.

Motor aspects: Adults over 50 may experience declining fine motor control which reduces their ability to manipulate small interfaces, including (multi) finger gestures. The inter-coordination of the eye and fingers to perform interactions is also reduced which affects the use of mouse, touchpad, and screen pointer devices. Screen targets for search interfaces should therefore be enlarged to allow older adults to accurately select these targets (*e.g.*, selecting a

The version of record is available at: http://dx.doi.org/10.1561/1500000073

search button or a recommendation for a query term). Distances between screen elements should further be increased and it is helpful to have a larger cursor. Complex mouse movements (*e.g.*, double-clicking, dragging, or even scrolling) should be reduced or avoided. The use of menus should be simplified (*e.g.*, by avoiding multi-level menu selections, leaving menus open until a selection has taken place) and keyboard use minimized. Furthermore, touch screen input should avoid multi-finger gestures, more time should be given for the user to complete operations, and tasks that require repetitive movements should be avoided to reduce strain.

Technological or attitude barriers: When designing search interfaces, older adults technical abilities and attitudes toward technology need to be considered. While older adults are not less willing to learn new technology, they often feel more anxious, less effective, and not as comfortable as younger people (Czaja and Lee, 2012). Usability, ease, and usefulness on the other hand lead to higher efficacy with technology. These barriers can be overcome by means of instructional design. (Czaja *et al.*, 2019, Chapter 8) offers guidelines in how to create training and learning resources, some of which are useful in overcoming barriers to search interface technology.

Despite abilities declining with age, a vast array of individual differences exist across the span of age brackets. This means that despite a downward trend in performance, many older adults demonstrate abilities that considerably outweigh those of younger adults (Czaja *et al.*, 2019, Chapter 3).

Touch-based interfaces are easy to learn and use by older adults without requiring them to have specific knowledge and experience with technology (Häikiö *et al.*, 2007) and offer faster pointing time than standard computer mice. These features level the playing field and older adults perform similarly to younger age groups (Murata and Iwase, 2005). When using smartphone applications, older users have issues with small-sized fonts and icons, lack of search features, and complex menu hierarchies (Gonçalves *et al.*, 2017). Alternatively, voice-based interfaces, often used in the emerging area of assisting older adults to

The version of record is available at: http://dx.doi.org/10.1561/1500000073

live well enough on their own (Portet *et al.*, 2013; Kowalski *et al.*, 2019; Shalini *et al.*, 2019) and to manage their health (Cheng *et al.*, 2018), are easier to use. The system by Cheng *et al.* (2018) combines the Google voice interface with a conversational agent to visualize health data for people with type-2 diabetes. Many of their users prefer the voice assistant over the text-based mobile phone counterpart. Kowalski *et al.* (2019) evaluated a voice assistant for older adults in a smartphone environment to better understand the benefits and barriers of voice interactions. Users valued the intuitiveness of the interface and the friendly and patient interaction style of the system. The interface helped them to gain independence by controlling devices that would otherwise require physical intervention. Barriers were the higher cost in time, the lack of sensors and screens in their home systems (*e.g.*, older TVs that could not be included in their context and could not provide feedback), the lack of device variety, and the fear of malfunction and system dependence. In a similar vein, the users described in Portet *et al.* (2013) also highlighted that voice assistants should allow for greater independence but only so long as the system did not encourage a lazy lifestyle with negative health benefits. Generally, the older users preferred voice assistants outright (Shalini *et al.*, 2019; Cordasco *et al.*, 2014), preferred them over a mobile phone application (Cheng *et al.*, 2018), and liked them for being easy to use, satisfactory, and for not requiring additional technical skills (Cordasco *et al.*, 2014). This suggests that many age-related issues originating from declining abilities can be compensated with such smart (search) interfaces.

4.4.3 Support for Disabilities

About 15.6 % of people worldwide (∼650 million) have a disability — 11.8% in higher-income and 18.0% in lower-income countries — with ∼ 2.2% suffering from a severe disability (WHO, 2011). Research on disability in the information sciences is generally focused on technology, viewed from the perspective of the content provider (*e.g.*, the library), limited to accessibility testing, and has had little or no involvement of disabled users in the studies (Hill, 2013). Additionally, there is also a strong focus on applications and an emphasis on visual disabilities.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

While many studies focus on visual impairments and dyslexia, there is less coverage on hearing disabilities, motor disabilities, autism, and down syndrome (Berget and MacFarlane, 2019) and more focus on information search than information seeking. In the following, we differentiate work within the two categories of perceptual and cognitive, and motor disabilities based on Berget and MacFarlane (2019).

Perceptual and cognitive aspects: Yoon *et al.* (2016) showed that many library websites are unsuitable for people with visual impairments who rely on screen readers. Sahib *et al.* (2012) observed both visually impaired and sighted users during complex information-seeking tasks to learn about the challenges of screen readers. They offer only limited support for query formulation (*e.g.*, query suggestions and spelling). Consequently, people with visual disabilities spend more time exploring search results and determining relevant pages while submitting fewer queries and visiting fewer result pages. They prefer recreating over reformulating queries and taking notes instead of re-issuing previous queries (Sahib *et al.*, 2012).

Some notable attempts have been made to remedy the situation. Auditory search result previews have been suggested as an exploratory aid to compensate for the shortcomings of screen readers. The work by Beinema (2017) shows a search system for visually disabled users that clusters search results to reduce the amount of result list processing. Sahib *et al.* (2012) further highlights the potential benefit of the integrated tracking of search results, and Gooda Sahib *et al.* (2015) present an example of a specific search interface for complex information seeking.

Berget (2015) investigates the database searching behavior of dyslexics with a rigid system that does not offer support for spelling and query building. They found that dyslexics needed more time to create more and shorter queries with more spelling errors and a high reliance on external help. This means that dyslexia negatively impacts search performance with systems that lack error-tolerance and support. Berget *et al.* (2016) studied how dyslexia affects the processing of search interfaces layouts (text

The version of record is available at: http://dx.doi.org/10.1561/1500000073

and icons) on the level of eye-tracking. They found that people with dyslexia need longer to get to interface targets with text. When looking at cognitive skills on query formulation – specifically decoding, short-term memory, and rapid automatized naming – they found that decoding skills correlate with query lengths and the number of errors made while short-term memory relates to the number of query iterations. Rapid automated naming had an impact on how long it took people to create queries.

Furthermore, Qu *et al.* (2019b) reviewed depression as a cause for memory impairments – specifically the effects of negative bias, and over-generalization – to inform user interface design for memory support. They suggest the use of memory banks for the accumulative recording and later for the selective retrieval of positive memories. Unlike lifelogging, the emphasis is on active curation, engagement, and management of these positive memories rather than large-scale and automated life-logging data collection and processing (Hoven *et al.*, 2012; Harvey *et al.*, 2016). Qu *et al.* (2019b) suggest that user interfaces should support more effortful, generative retrieval of memories and further recommend that interfaces should actively support users in re-experiencing positive memories.

Vtyurina *et al.* (2019) found that although screen readers allow visually disabled people to more comprehensively engage with content, voice-based assistants offer accessibility, they prove to be less suitable for creating an overview or for engaging more deeply with content. The researchers then combined both technologies in a search interface prototype – an example of how multiple 'imperfect' approaches may be used to better support completing tasks for those with a disability.

Motor: Berget and Sandnes (2019) note that the recognition of motor disabilities for information seeking and search is mostly limited to guidelines and policies (*e.g.*, general interface design recommendations, and advice for libraries). Although not specific to search interfaces, Gajos *et al.* (2010) present an AI-supported system that generates user interfaces based on a formal description of

The version of record is available at: http://dx.doi.org/10.1561/1500000073

people's motor abilities and limitations, a cost function, and a usage model of the interface. Such a system could be applied to either generate, assist, or evaluate search interfaces for their suitability for specific user profiles. While there is little specific work, a recent review evaluates information seeking and search models for how well they are suited to studying users with disabilities (including motor disabilities) (Berget *et al.*, 2021). This may be a useful step in further enhancing research in this direction.

4.5 Summary

This section has reviewed the research coming from interdisciplinary perspectives on adaptive user interfaces with a special consideration for specific user groups, notably children, older adults, and people with disabilities. Such interfaces can adjust to the user in response to signals picked up by sensors that lead to system adaptations. We have discussed the design problem these groups pose for the design of search interfaces and the current state of design solutions.

- Adaptive user interfaces are designed to cater to user tasks, performance, intentions, goals, and user characteristics. The sources of signals for inferring the user states include user search behavior, physiological signals, the device platform, and the contextual environment.
- The personalization of user interfaces can be accomplished through adaptation to user models of intentions, goals, and tasks by search behavior, eye gaze, and multiple sources of physiological signals. Personalization by personal attributes can provide system-driven customizations to support user interactions. User perception interaction data together with eye gaze data or physiological signals can be used to predict the user's emotional states.
- Contextualization of user interface research examines the relationships among situational context, search behavior, and user preference as well as collaborative information seeking and retrieval.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

- Search interfaces designed for children consider their cognitive development, reduced capacity for fine-motor control, and emotional support. Search interfaces for older adults specifically consider diminished cognitive abilities like attention and memory, physical abilities, and attitudes toward technology. Search interfaces to support people with disabilities consider sensory, cognitive, and motor abilities.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

5

Evaluation of Search Interfaces

This section aims to help readers be mindful of approaches and measures and the reasoning for choosing different approaches and measures in the evaluation of search interfaces. This section summarizes the evaluation approaches and measures used in the empirical evaluation research papers reviewed in this monograph. The evaluation methods for search interfaces have borrowed methods from HCI, social science, traditional IR, and interactive IR. As there is no ideal method that fits all kinds of evaluation purposes, we provide a road map for researchers and practitioners in selecting appropriate methods for specific purposes of evaluation.

In this section, we classify evaluation studies according to the evaluation period of the product design process (formative evaluation and summative evaluation), and according to the specific goals of research (exploratory, descriptive or explanatory), and then discuss a series of evaluation approaches, and what kind of results one could expect from each evaluation approach. More importantly, we argue that for the evaluation of search interfaces, one should first refer to the theoretical models or frameworks (as discussed in Section 2), consider which components one intends to evaluate (as discussed in Section 3), and

The version of record is available at: http://dx.doi.org/10.1561/1500000073

consider appropriate evaluation techniques for the current evaluation. In addition, this section classifies evaluation measures along two dimensions, evaluation objects and evaluation assessors, and discusses the relationship among different types of evaluation measures.

5.1 Formative Evaluation and Summative Evaluation

The evaluation of search interfaces could take place at any point in the product life cycle. For example, if a new search system is to be designed, considerable time would usually be devoted to discovering users' requirements and expectations; when these requirements have been established, they are used to create initial sketches or prototypes, and the evaluation at this stage is to assess whether the design has embodied users' requirements appropriately; if it is an upgrade of an existing search interface, then the main goal is to ascertain what needs to be improved. Evaluations which check during the design process whether the product continues to meet users' needs and requirements are called formative evaluations whereas the evaluations carried out to assess the success of a finished design are known as summative evaluations.

Considering the period of evaluation during the product design process, there are mainly two types of evaluations: formative evaluation and summative evaluation (Sharp *et al.*, 2019). Formative evaluations are conducted when a new type of search system is proposed or during the design to check that an information system meets users' needs, for example, before or during the design of a conversational search system. Summative evaluations are carried out to assess the success of a finished product or when the product is to be upgraded. Most evaluation research studies are concerned with evaluating certain parts of the search interface, using either the formative or the summative evaluation. However, there are also a few research studies that have comprehensively presented the formative evaluation process when designing a new search interface. Among the empirical studies appearing in this review book, Schlötterer *et al.* (2020), Ye *et al.* (2020) and Druin (2005) are selective studies that perform formative evaluations during the design process.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Schlötterer *et al.* (2020) present the formative evaluation they carried out with non-expert users when designing QueryCrumbs, a compact and easy-to-understand visualization for navigating the search query history, supported with iterative query refinement. They applied a multi-layered interface design to support all users, including novices and first-time users as well as intermediate and expert users. The visualization was evaluated with novice users in a formative user study and with experts using think-aloud protocols. Its usage was also evaluated in a long-term study with software logging. The formative evaluation with novice users showed that the interactions could be easily performed and that the visual encoding was well understood without instructions. Results indicate that QueryCrumbs can support users when searching for information in an iterative manner. The evaluation with experts showed that expert users could gain valuable insights into the back-end search engine by identifying specific patterns in the visualization. In a long-term usage study, uptake of the visualization was observed, indicating that users deemed this new visualization system beneficial for their search interactions.

Ye *et al.* (2020) reported on a two-year project in which they applied design principles to develop effective and usable visualization solutions for combustion scientists. Their process consisted of three stages: the user and task analysis stage, the iterative design stage, and lastly the full realization stage. At the first stage, they worked closely with combustion scientists to understand what they truly desired; at the second stage, they made several important design decisions and built a complete working prototype for testing; at the last stage, they conducted formal evaluations of the design. They conducted user experience tests to identify and correct unforeseen problems or issues. This study, which carefully records their design process, demonstrates the importance of user-centered design principles and describes lessons learned over the design process.

Druin (2005) reviewed her research process in designing digital libraries for children. In her formative evaluation, she involved children as technology design partners. In brainstorming with them about how children search for books, she found that the techniques they used included sketching new ideas with art (low-tech prototyping) and critiquing

The version of record is available at: http://dx.doi.org/10.1561/1500000073

existing technologies with the use of post-it notes. She suggests that children can contribute important information and, therefore, should be involved in the design of digital libraries for children.

A common practice in evaluation is to involve real target users at the beginning of the formative evaluation. Different groups of users may require different communication skills. As we have shown above, brainstorming may work best with children (Druin, 2005), while designers may need to work closely and have discussions with scientists to understand what they truly want (Ye *et al.*, 2020). We may need different communication methods with novice and expert users, as demonstrated by Schlötterer *et al.* (2020). Since an understanding of users' requirements and expectations learned by self-reported data such as questionnaires or interviews may not be reliable, triangulation of findings from different types of evaluation is needed to obtain a full picture.

5.2 Evaluation Objectives

We can also consider evaluation purposes in terms of the specific goals of research: exploratory, descriptive, or explanatory. Such characterization is commonly discussed in research methods textbooks but not often discussed in the evaluation approaches of search interfaces. However, when conducting different evaluation studies, we can also ask similar questions concerning the evaluation objectives:

- Is it an exploration of users' expectations of a search interface in specific contexts or by a particular group of users?
- Is it a description of how users are using an existing or a newly developed search interface?
- Does it examine the relationship between two or more variables with the goal of explanation and prediction?

We do not consider exploration an evaluation study since the main goal is to understand what users need or want to do in certain contexts. However, we should keep in mind that the main criteria of evaluating a search interface are whether the design meets the user's needs and how

The version of record is available at: http://dx.doi.org/10.1561/1500000073

well the interface facilitates task completion. Therefore, understanding user needs is often the first step of all the evaluation exercises.

Research objectives often include three main categories: exploratory, descriptive, and explanatory in social science research methods (Creswell and Creswell, 2018; Williamson, 2018). We adopted the categories of this branch of science to classify the objectives of studies evaluating search interfaces. In exploratory evaluation studies, the main goal is to explore users' needs, requirements, or expectations for a certain type of search system or in certain contexts. The main goal for descriptive evaluation studies is to describe how users are using the current, or newly developed, search interface, or to evaluate their search experience, preferences, or performance. The distinction between an exploratory evaluation study and a descriptive study is that at the stage of the descriptive study, researchers already know users' general expectations in those circumstances and already have the interface or prototype developed so that they can evaluate how users are using the interface.

Concerning explanatory evaluation studies, the main goal is to improve users' search experience by examining factors that could influence users' search behaviors or search experience. Explanatory evaluation studies are often in the form of controlled experiments which aim to advance the field's understanding of how people use a certain type of search interface to determine which design concepts work well under what circumstances and why. The explanatory evaluation studies may compare two or more competing interfaces to help decide if a new feature or a change could improve the performance of an existing interface. It is possible that some research projects have multiple research goals and include exploratory and descriptive or explanatory studies in one research paper. The relationship and differences among the three evaluation objectives are shown in Table 5.1.

5.2.1 Exploratory Evaluation Studies

Exploratory evaluation studies are often conducted before a new search system is designed for a particular group of users, for instance, search interfaces for children to select books (Druin, 2005) or for elderly people to read and browse (Piper *et al.*, 2017), or when designing a new type

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Table 5.1: Classification of evaluation objectives

Evaluation Objectives	The Main Goal	Evaluation Stage
Exploratory evaluation	to explore users' needs, requirements, or expectations for a certain type of search system, or in certain contexts	before a new search system is designed
Descriptive evaluation	to describe how users are using the current, or newly developed search interface, or to evaluate their search experience, preferences, or performance	after the prototype or the product is developed
Explanatory evaluation	to improve users' search experience by examining factors that could influence users' search behaviours or search experience	not necessarily focused in a specified search system

of search system, for example, search interfaces for leisure search or search interfaces with AR or VR (as reviewed in Section 6). The main goal is to help researchers and practitioners to understand users and their needs in certain search circumstances, to identify opportunities for new technology to be implemented in the search interfaces. It is also possible that through exploratory studies, researchers can propose new design concepts or new features from the findings of these studies.

The evaluation methods applied in exploratory studies are primarily qualitative, such as field studies combined with interviews or focus groups. When the target users are hard to find or difficult to observe, case studies with 3-5 participants or fewer would also be helpful. Some studies may also start by analyzing online logs of user-generated content to get an overview of users' expectations. For example, Bi *et al.* (2019) adopted a mixed-method approach "to build an in-depth understanding of the experience and corresponding needs for the interaction of R &S [runners and spectators] during LDRE [long-distance running events]" (Bi *et al.*, 2019, p. 3). They first analyzed online blogs of marathon runners and their reviews of related apps and then conducted

The version of record is available at: http://dx.doi.org/10.1561/1500000073

semi-structured interviews and observations, and finally, they provide design implications for technology connecting runners and spectators. It is a comprehensive exploratory study of shared experiences of marathon runners and spectators. A survey may not be an appropriate method for exploratory studies, since it could not provide what users would do in real settings and it is difficult to elicit users' needs or expectations from questionnaires.

Exploratory studies usually are followed by more in-depth research, (*e.g.*, descriptive research). The results of exploratory studies are usually not aimed to be specific or comprehensive, but they could help researchers better understand user needs, users' contexts, the new technology, and their interactions. Researchers will be able to propose further research questions or design principles or to reduce the research scope based on the findings.

5.2.2 Descriptive Evaluation Studies

The descriptive evaluation studies are conducted after the prototype or the product is developed, so the evaluation would describe how users are using the newly developed search interfaces, or to evaluate the usability of the interface or users' search experience, preferences, or performance. In general, descriptive research aims to answer 'what it is' or 'how it is' type of questions. In the descriptive evaluation research, researchers focus on documenting and describing how users are using the newly developed search system, providing concrete and nuanced user interactions or profiles, to identify design defects.

The evaluation methods that can be employed in descriptive studies include traditional HCI evaluation techniques, (*e.g.*, heuristic evaluation, cognitive walkthrough, usability tests in laboratory settings, and field observations in naturalistic settings). For example, Guy (2018) conducted data mining on search logs of voice search systems to characterize users' voice search behaviors. Aliannejadi *et al.* (2019a) conducted field observation studies to describe users' search tasks at different times in a day or a week and how the surroundings or transportation tools would influence users' search goals or experience. Descriptive evaluation studies, particularly the published descriptive studies, are often compre-

The version of record is available at: http://dx.doi.org/10.1561/1500000073

hensive and specified, focusing on certain types of contexts or search systems. The expert-based evaluation methods (*e.g.*, heuristic evaluation or cognitive walkthrough) and usability tests are often conducted to guide the industry practice in designing search interfaces.

5.2.3 Explanatory Evaluation Studies

The main goal of explanatory evaluation studies is to explain why certain phenomena occur and why users may prefer a certain type of interface design. Since this type of evaluation study is often concerned with the causal relationship among the variables, the evaluation may not necessarily focus on a specified search system. Explanatory studies often use more structural or focused methods than exploratory or descriptive studies, (*e.g.*, controlled experiments) and propose research hypotheses for testing rather than take the format of research questions. As Kelly (2009) explained, explanatory studies may not always offer an explanation and studies often make predictions of user interactions but stop short of pursuing any explanation. However, their explanations are the foundations of the theoretical development of user behavior and design methodologies. Therefore, even though the explanatory evaluation studies may be designed to evaluate a certain search interface, they could offer design guidelines for a certain context or for a given group of users.

The work of Harvey and Pointon (2017) is a good example of an explanatory evaluation study. The authors examined two factors of user performance and perception: mobile situations and device type. They conducted a lab experiment with 24 participants each of whom were given one of two device types: a tablet or a phone. The mobile situations were designed according to the distraction level: walking quickly on a treadmill, navigating an environment with obstacles, and a baseline condition in which the participant was seated without any distractions. The results showed the simulated conditions significantly affected both participants' objective and perceived search performance; the type of device also impacted how users felt about the search tasks, how well they performed, and the amount of time they spent engaged in the tasks. This explanatory evaluation study helps us better understand

The version of record is available at: http://dx.doi.org/10.1561/1500000073

how context and device size affect search behavior and user experience and provides insight to inform the design of future interfaces for mobile search.

Wittek *et al.* (2016) examined information-seeking in an uncertain environment such as user interactions with experimental search interfaces. This study investigated visual search behavior in information seeking by taking measurements of risk (hesitation) and ambiguity (opportunity cost). Since information seeking was conceptualized as an information foraging scenario comprising sequential decision making, this study was an attempt to understand why users with different cognitive styles use different search strategies under an uncertain environment. This is an example of an explanatory evaluation study conceptualized and informed by the information foraging theory (see Fu, 2020, for more examples).

5.3 Types of Evaluation

After deciding on the general evaluation objectives, researchers need to select appropriate evaluation methods to address the objectives. In making that selection, researchers need to consider both the design stage of the search interface and the evaluation objective. Different types of evaluation techniques can be applied, depending on the type of product, the prototype or design concept, and the value of the evaluation to the designers, developers, and users. Key aspects in choosing the evaluation methods lie in making decisions as to whether real users will be involved in the evaluations and whether the evaluation will be conducted in controlled settings or not. Several HCI evaluation techniques, such as cognitive walkthrough and usability testing, are expert-based evaluations without real users, and carried out with limited resources. We recommend that evaluation studies of search interfaces include real users since user preference and performance are specifically considered. In this section, we categorize the evaluation approaches of search interfaces into three categories: 1) evaluations without real users; 2) evaluations with real users in realistic settings; and 3) evaluations with real users in controlled settings. We then explain the main consideration for each type of evaluation method.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

5.3.1 Evaluations Not Involving Real Users

Evaluations of search user interfaces often involve real users, but sometimes do not. The evaluations that do not involve real users are usually expert-based methods, for example, heuristic evaluation and cognitive walkthrough. These are two well-known expert-based methods that can identify a large number of problems using small amounts of time and financial resources. In these methods, three to five users are considered sufficient to perform evaluations (Nielsen and Landauer, 1993). It is common to implement this type of evaluation during the developmental stage of the search interface design. As a typical method in HCI, this approach does not involve the recruitment of users with their own information needs and it can be relatively efficient in detecting design problems. However, it does not mean these approaches are easy to do. For example, cognitive walkthrough requires researchers to fully understand users' typical tasks, goals, and sub-goals for each task when using a given search system. As we have argued in Section 2, tasks should be fully examined and clarified in realistic contexts before implementing cognitive walkthrough evaluations to improve the validity of the study. The cognitive walkthrough is commonly used in industry, but it seldom appears in research publications.

Heuristic Evaluation

Heuristic evaluation is guided by heuristic principles to identify user interface designs that violate these principles. Heuristic evaluation is done by expert evaluators examining the design of a user interface and judging its compliance with a list of predefined principles (heuristics). These principles are used as a template to help the evaluators identify the potential problems users may encounter. One of the problems of using heuristics is that designers can sometimes be led astray by findings that are not as accurate as they first appeared to be[1].This problem can arise from different sources such as from a lack of experience, and from the biases of the user experience researchers who conduct the heuristics evaluations (Sharp *et al.*, 2019).

[1] http://www.usefulusability.com

The version of record is available at: http://dx.doi.org/10.1561/1500000073

One of the evaluation studies we review in this monograph is a voice interface application using the Google Home for elderly patients with Type-2 diabetes which Cheng *et al.* (2018) developed and then evaluated for the usability of the interface. The researchers combined three usability evaluation methods, first a feature-based comparison of this application against other available mobile apps, a survey to measure user satisfaction with the application, and lastly the feedback and evaluation from elder-care experts and potential users. The comments from the experts helped determine the shortcomings and possible solutions that could enhance the usability of the interface.

Cognitive walkthroughs

A cognitive walkthrough is a task-oriented and structured method of evaluation that focuses on evaluating user interface designs for ease of learning (Sharp *et al.*, 2019). To perform the cognitive walkthrough evaluation, researchers or practitioners must first determine initial user goals and sub-goals based on each scenario and make an action sequence list. When designing cognitive walkthrough evaluations, it is recommended that an adequate number of scenarios be used to ensure all users' tasks are covered. Independent evaluators then examine how easy it is for new users to accomplish tasks with the system. They go through the user interface by exploring every step needed to complete each task noting user goals, user sub-goals, user actions, system responses, and potential user interaction problems. Each evaluator then independently provides a list of usability problems.

Cognitive walkthroughs could be considered an expert-based evaluation method, and usually no real users are involved in the evaluation. Some recent studies in IR employ a simulated user method to assess the search performance of the system. For example, Koskela *et al.* (2018) simulated a setting where a user is writing a text about a given topic and then simulated user interactions in two types of search tasks, an exploratory search task and a known-item search task, to evaluate the performance of the proposed search intent prediction in a proactive IR system. The user simulation method is similar to a cognitive walkthrough in that neither involves real users; however, the difference is

The version of record is available at: http://dx.doi.org/10.1561/1500000073

that the user simulation method can only assess objective performance. It is not suitable for judging the subjective usability of the system.

Khajouei *et al.* (2017) compared two expert-based evaluation methods, heuristic evaluation and cognitive walkthroughs, in terms of the number of identified usability problems, their severity, and the coverage of each method. The results showed no significant difference between the two methods in terms of the total number of identified usability problems. However, cognitive walkthroughs worked significantly better for identifying usability problems that affect the learnability of the system, and heuristic evaluation performed better for detecting problems that result in user dissatisfaction. These results infer that cognitive walkthroughs would be the preferred method for evaluating systems intended for novice users while heuristic evaluations would be better for users who have experience with similar systems.

Analytics

Analytics is a method for evaluating user traffic during a period when an information system is in use. Web analytics, which could be collected locally on users' devices or remotely across the Internet, are used to examine traffic on a website or part of a website. Analyzing logged user interaction data could help the researcher understand what part of the website has been used and when. For example, Jiang *et al.* (2017) analyzed the transaction log file from a Chinese university library's online public access catalog (OPAC) using a clickstream data analysis framework. The results showed that users relied heavily on the single-box simple search interface, seldom involved themselves in an exploratory search process, and preferred page navigation over search refinement when interacting with search results. It indicated that the OPAC was used as a lookup tool to locate known academic resources, rather than a discovery tool. Likewise, after Schlötterer *et al.* (2020) designed QueryCrumbs, the compact visualization interface for navigating the search query history, they used analytics to examine the long-term usage of the visualization tool. It revealed that there had been an uptake in the use of this new visualization system, indicating that users found QueryCrumbs beneficial for their search interactions.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

5.3.2 Realistic Settings Involving Real Users

Evaluation studies in natural settings with real users are often conducted with either little or no control imposed on participants' activities. These studies are often called field studies. Field studies are used primarily to help identify opportunities for new technology, establish the requirements for a new design, and facilitate the introduction of technology or inform the deployment of existing technology in new contexts. Another reason to use field studies is that technologies are being developed for use outside office settings, such as in the home, outdoors, and in public places.

Typical data collection methods for field studies include observation, interview, questionnaire, focus group, diaries, and sometimes log analysis. In these studies, qualitative accounts and descriptions of people's behavior and activities are obtained that reveal how they used the product and reacted to its design. The results of these studies may not directly provide suggestions for the specific design of the search interface, but they are very important in informing researchers and developers' understanding of users expectation of various user groups and in different contexts. These methods are often conducted at the early stage of the interface design to provide general suggestions for the components of the search interface that may need to be further developed or refined.

When investigating what children need and want for accessing digital libraries, most studies asked experts, usually adults, to talk about children; however, Druin (2005) had her research group work together with children and brainstorm to better understand their needs. Children are a special group; they follow adults' directions because of the existing power structures. To encourage the children to express their thoughts freely and confidently, Druin insisted on three basic practices — no raising hands, using first names only (no last names or titles), and wearing informal clothing.

Piper *et al.* (2017) conducted in-depth interviews with 15 older low vision or blind adults to understand how they use technology to communicate and seek information. The interviews were conducted in their homes or in a private room within their residential community.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

This was an exploratory evaluation and the main goal was to detect the challenges they face and to see these as opportunities to design new technologies so that this demographic will be able to participate in the information society more fully.

5.3.3 Controlled Settings Involving Real Users

Many of the evaluation studies are conducted in controlled settings, (*e.g.*, usability testing and controlled experiments).

Usability Testing

Usability testing has traditionally been done in controlled laboratory settings to evaluate whether an interface being developed is usable by the intended user population for successful task completion. Usability testing in a controlled lab enables evaluators to control what users do and to control environmental and social influences that might impact the users' performance. The ultimate goal of usability testing is to identify usability problems, that is to test whether the interface being developed is usable by the intended user population to accomplish the predefined tasks. When conducting usability testing, researchers often collect data about users' performance on predefined tasks, users' interactions (*e.g.*, keystrokes and mouse movements), and users' self-report measures about their satisfaction with the product. Sometimes participants are asked to think aloud while carrying out tasks, or interviews (structured or semi-structured) may also be conducted to collect additional information about why participants liked or did not like the interface.

Bickmore *et al.* (2018) conducted a usability test to evaluate the performance of three conversational assistants (Siri, Google Assistant, and Alexa) when asked nontrivial questions about medical situations. A total of 54 subjects participated in this evaluation study. Each conducted three medication tasks with each conversational assistant, a total of nine tasks. After completing the third task with a given conversational assistant, the subject filled out a satisfaction questionnaire. After completing the interactions with all three conversational assistants, subjects were interviewed about their experience. The measures selected in this study include health literacy, computer and conversational assistant

The version of record is available at: http://dx.doi.org/10.1561/1500000073

literacy (self-report measures), satisfaction measure (self-report measures), time per task and time per attempt (objective measures), and task performance measures (*i.e.*, task failure, potential resultant harm, and potential resultant death) using a professional scale adopted from medical domains. The results of this usability test revealed that, in 2017, conversational assistants failed more than half of the time and led subjects to take actions that could have resulted in harm.

Most usability testing used self-report measures in questionnaires; however, Murcia-López *et al.* (2020) argue that the scales in questionnaires are difficult to interpret. For example, one participant's score of '5' out of a maximum of '7' might mean something completely different than another's. They further argue that physiological measures do not provide a universal solution since they are complex and have limited utility. Scale development and validation are critical for self-report measures. Slater *et al.* (2010) introduced a method based on an analogy with colorimetry that potentially overcomes such methodological challenges. Instead of asking participants to judge how 'red' a color is, they allowed participants to manipulate the color projectors to match their perception of color by adjusting red, green, and blue projectors. Likewise, when evaluating Virtual Reality experiences with a virtual human character, Murcia-López *et al.* (2020) allowed participants to spend a virtual budget to modify factors to incrementally improve their quality of experience. Participants could stop tweaking the factors when they felt further changes would not make any further difference. Through this method, evaluators allowed the users to calibrate their preferences, and from that, evaluated which part users want to change.

Controlled Experiments

Most of the time, when we talk about controlled experiments, we refer to the experiments conducted in labs, but controlled experiments can also be done in natural settings. Recently, there have been new types of experiments, (*e.g.*, the Wizard of Oz study, and crowdsourcing). These evaluation methods are described and discussed in this section.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Experiments The experiment is the most common approach used in the empirical studies examined in this monograph. The experimental design strictly controlled most of the variables to investigate the relationship between independent and dependent variables. Most of these experiments aimed to examine how the factors (*e.g.*, task type, user type) affect users' preferences on different design options for the search interfaces. Most of the published IIR studies have used user experiments as the main approach for evaluation research. Sometimes they labeled their approaches usability tests, (*e.g.*, Karaman and Sezgin, 2018), or observational user studies, (*e.g.*, Sahib *et al.*, 2012), but since they specified the independent and dependent variables, and since their goal was to examine the causal relationship between the variables, they should be classified as experiments.

In experimental design, several issues may influence the validity of the findings of the research. For the scenarios in the experiments, it is recommended that the researchers construct simulated work/search tasks for participants to conduct searches for evaluation (Borlund, 2013). Besides, when there is more than one search interface to be evaluated or more than one task to be completed during the evaluation, the order of the search interfaces or tasks should be rotated to reduce the possibility of an order effect; for example, a Latin square is an effective way to counterbalance the order and reduce the number of participants for full rotation (Kelly, 2009).

Experiments can be conducted in natural settings or remotely. For example, Sahib *et al.* (2012) carried out their observation of participants remotely using Skype via the screen-sharing functionality. This is also common in usability testing in HCI and has been found to be as effective as traditional settings.

Wizard of Oz Study Wizard of Oz studies receive their name from the well-known book by the same title. These studies are similar to the Wizard of Oz from the book in that researchers often imitate 'grand' systems that they would like to study. Wizard of Oz studies can be used for proof-of-concept and to indicate users' expectations and possible interactions with the search interfaces. This simulation method is commonly used in the design of a conversational search system. For

The version of record is available at: http://dx.doi.org/10.1561/1500000073

example, Cordasco *et al.* (2014) evaluated the prototype of the vAssist system (a voice-controlled assistive care and communication service for the home) using a Wizard-of-Oz paradigm. They asked 43 elderly Italians to interact with the interface in 4-5 pre-defined scenarios and administered three questionnaires to measure their perception of the system's usability, learnability, and intuitivity. The results demonstrated the interface was greatly appreciated because of the simplification it provided the elderly in the everyday use of technological products.

Crowdsourcing Recently, many researchers have relied on Amazon's Mechanical Turk and other crowdsourcing platforms to recruit participants to evaluate search systems, assess user preferences, or make relevance assessments via the Web, (*e.g.*, Setlur *et al.*, 2020; Eickhoff, 2018; La Barbera *et al.*, 2020). The advantage of crowdsourcing is that it may provide researchers with access to a diverse pool of potential participants in a very timely and cost-efficient way. Crowdsourcing participants with self-administered data were found to be an alternative to traditional pretesting methods (Kelly, 2009). However, if the goal is to evaluate the performance or outcome of search systems, validity and reliability cannot be guaranteed during the process.

5.4 Evaluation Measures

Besides evaluation approaches, evaluation measurement is also fundamental to the evaluation of search interfaces. A large number of measures have been proposed to evaluate search systems, especially the algorithms of IR systems, but for the evaluation of search interfaces, standard measures are lacking.

Kelly (2009) classified all measures in IIR into four categories: contextual, interaction, performance, and usability. Among them, contextual measures are measures that characterize subjects or their information-seeking situations, and the other three categories of measures could all be considered evaluation measures. The interaction measures could either be whole-session measures describing the total effort (*i.e.*, number of queries, number of search results viewed, number of documents viewed, number of documents saved) that can only be obtained at the end of search sessions, or behavioral measures that could be detected and calcu-

The version of record is available at: http://dx.doi.org/10.1561/1500000073

lated during the search process (Liu *et al.*, 2014). White (2016) argued that there has been less attention paid to process-oriented measures in IR evaluationeven though they provide useful information into the nature of the search process. He further classified process-oriented measures into seven aspects: learning, efficiency, cognitive load, serendipity, enjoyment/happiness/pleasure, frustration, and engagement.

Wilson (2011) highlighted two types of evaluation measures of search interfaces. One is search-based measures which are mainly obtained from the query log, eye movements, or mouse movements on search interfaces to understand how different search interface features are used during search; The other is subjective feedback from subjects, such as search difficulty, usefulness, engagement, emotion, cognitive load, search, and learning performance. Besides these common subjective measures, the choice of measure should also vary significantly depending on the type of IR system. For example, for the evaluation of a conversational search interface, Kocaballi *et al.* (2019) analyzed the assessment items listed in the six main questionnaires and coded them into one overall measure and eight different user experience dimensions: Affect/Emotion, Enjoyment/Fun, Aesthetics/Appeal, Hedonic/Quality, Engagement/Flow, Motivation, Enchantment, and Frustration.

After examining the empirical studies reviewed in this monograph and combining all the measures mentioned in the above reviews, this section summarizes the measures as two main dimensions in the evaluation of search interfaces: on one dimension are the evaluation objects, which could either be process or outcome, and on the other are the evaluation assessors, which could be objective or self-reported measures (as depicted in Table 5.2).

The objective-process quadrant mainly includes behavioral traces in interaction logs, as well as physiological signals captured using different sensing equipment (*e.g.*, eye trackers, EEG, fMRI). These behavioral measures may consist of frequency counts of the activities that occurred, which can be directly related to interface functionality (Kelly, 2009), the vocabulary used in query statements as representative of learning (White, 2016), changes in facial expression interpreted as happiness (Moshfeghi and Jose, 2013), or modification of concept maps or mind maps as representative of learning structural changes (Liu *et al.*,

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Table 5.2: Classification of evaluation measures

Objects / Evaluators	**process**	**outcome**
objective	behavioural traces in interaction logs	search performance: precision, recall, MAP, NDCG, etc.
self-reported	self-reported process or search experience measure, e.g. via think-aloud or simulated recall, etc.	self-reported overall search success, satisfaction, learning performance, etc.

2020a). The strength of these objective-process metrics is that they can also be considered indicators of implicit relevance feedback, users' preferences, or user's self-reported measures. However, the interpretation of these interaction measures is challenging and needs careful consideration and comparison (Kelly, 2009); for example, Hassan *et al.* (2014) disambiguated some long search sessions as being indicative of the user exploring, not struggling. In addition, not all behavioral metrics are good indicators of the search process. For task-based evaluation, the behavioral measures that can be captured or calculated during the search process are more important than the total account of behavioral measures since they reflect the user's instant search state and could help generate predictive models (Liu *et al.*, 2012).

The self-reported-process quadrant mainly represents self-reported measures about the search process or search experience via think-aloud during the search process or via simulated recall or task interview after the search task is complete. Evaluators usually ask participants to recall their cognitive load (*e.g.*, the NASA Task Load Index (Na, 2021), frustration (Feild *et al.*, 2010), difficulty (Liu *et al.*, 2014) or engagement (O'Brien *et al.*, 2020)), and then construct predictive models of these subjective measures based on users' behavioral measures.

The objective-outcome quadrant includes measures that evaluate search performance (*e.g.*, precision, recall, MAP, NDCG, etc.) (White, 2016; Kelly, 2009) and users' knowledge (through an assessment of the users' knowledge level (Zhang and Liu, 2020; Bhattacharya and Gwizdka, 2019) or through an assessment of the quality of learning outcomes (Wilson and Wilson, 2013)).

The version of record is available at: http://dx.doi.org/10.1561/1500000073

The self-reported-outcome quadrant represents users' subjective evaluation of the search outcome (*e.g.*, search success (Vakkari *et al.*, 2018; Spiller *et al.*, 2021)), learning outcome (learning performance, or knowledge level (Collins-Thompson *et al.*, 2016; Capra *et al.*, 2018)), or their overall experience of the search interface (satisfaction, easiness, enjoyment, engagement, and appeal of use (Cordasco *et al.*, 2014)).

Traditional evaluation of search systems focuses on objective-outcome measures such as search performance, and research in HCI stresses the self-reported evaluation of search success and satisfaction. These are both outcome-oriented measures. However, as more diverse search systems appear and searching becomes more interactive and iterative, more process-oriented measures should be considered in the evaluation of search systems, especially search interfaces. In addition, it is also important to distinguish the objective measures from self-reported measures. The self-reported measures are users' assessments of their search experience during the search process, but these measures cannot be easily observed by the system. Research has been working on generating the relationship between behavioral traces in interaction logs with users' self-reported process or experience measures. When generating predictive models, it is important to consider contextual factors as mediating variables. Process-oriented self-reported measures could also provide the utility of search interface features at different stages of the search process (Huurdeman *et al.*, 2016; Huurdeman, 2017). Future measures should also consider self-reported meta-cognition during the search process to help users have more control of their search process (Smith and Rieh, 2019).

5.5 Evaluating Search Interfaces

As introduced in this section, it is important to consider the theoretical models and frameworks, select the components of search interfaces, and apply appropriate evaluation techniques for evaluating search interfaces. We now illustrate these points by the evaluation of conversational search systems. Since research on conversational search systems is multidisciplinary and involves the fields of IR, natural language processing (NLP), and machine learning, there have been diverse approaches and evalua-

The version of record is available at: http://dx.doi.org/10.1561/1500000073

tion frameworks. One of the key distinctions is the intrinsic and extrinsic evaluation of machine translation outputs from the NLP community (*e.g.*, Clark (2015) and Sparck Jones (1994)). The intrinsic evaluation focuses on the internal outputs from the system whereas the extrinsic evaluation is concerned with how the use of the system contributes to external outputs, such as task completion. In the evaluation of NLP systems, Sparck Jones (1994, p. 103) advocated that "... establishing their merits by *intrinsic evaluations is of limited value, so extensive* extrinsic evaluation in a variety of environments is required." Belz (2009) proposed a combination of intrinsic evaluation techniques and extrinsic validation after considering the issues of 1) extrinsic meta-evaluation of evaluation metrics; 2) extrinsic evaluation of human-produced reference material; and 3) extrinsic evaluation of training data.

The Alexa Prize Socialbot Grand Challenge exemplifies the intrinsic approach to evaluation that belongs to descriptive evaluation studies without involving real users. It was designed as a research competition to advance our understanding of human interactions with socialbots, with the support of large amounts of user data from Amazon.com. Since the ultimate goal is to enhance the user experience, specifically user satisfaction when interacting with Alexa, it is not surprising that user satisfaction has been selected as the main evaluation criterion for success. Dinan *et al.* (2020) provide details about the evaluation criteria which consist of automatic metrics from the system and human evaluation sourced through Amazon's Mechanical Turk. Since the objective was to judge the system's performance based on approximations of user satisfaction, automatic metrics and human evaluation results were compared; however, a discrepancy was found between them.

In the IR community, evaluation efforts have focused on the creation of a test collection to be used in comparing system performance by appropriate evaluation metrics, or to compare the conditions of different types of information-seeking tasks under which search interfaces are used. For example, Shiga *et al.* (2017) attempted to model the information needs in collaborative search conversations for facilitating the evaluation effort in the development of conversational search applications. Avula *et al.* (2019) assessed the effect of using searchbots on user perceptions, experience, and search behaviors during information-seeking tasks to

The version of record is available at: http://dx.doi.org/10.1561/1500000073

support collaborative search. By contrast, Liu *et al.* (2020b) evaluated conversational search systems designed for pilots involving real users (*i.e.*, flight school students in a flight simulator). While each study is informative in and of itself, an extensive extrinsic evaluation of conversational search systems in various environments is needed to ensure that research aligns with the evaluation objective of fully supporting users' task completion.

5.6 Summary

This section reviewed the following aspects of the evaluation of search interface studies:

- The evaluation of the search interface at different stages of product design should consider different types of evaluations, namely formative evaluation and summative evaluation.
- Before the evaluation, it is also important to determine the general evaluation purposes: whether it belongs to exploratory evaluation, descriptive evaluation, or explanatory evaluation.
- The evaluation approaches of search interface were categorized into three types: evaluations without real users; evaluations with real users in realistic settings, and evaluations with real users in controlled settings.
- Concerning the evaluation measures, this section classified them along two dimensions: 1) evaluation objects: process or outcome; 2) evaluation assessors: objective or self-reported. The evaluation measures used in different quadrants reflect different aspects of the usage of the search interface.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

6

Emerging Trends

This section addresses selective and emerging topics in the field of search interfaces. Conversational search interfaces have grown substantially in importance, triggered by the rising popularity of chatbots and intelligent personal assistants through progress made in the development of the smart Internet of Things (IoT) devices. Leisure-related search has gained momentum over the last decade facilitated by search engines that are becoming increasingly accessible and relevant for our private life. Search interfaces that specifically support the discovery of new and unintended information by supporting serendipity and creativity are yet another recent development. Finally, we highlight how search interfaces for Augmented, Mixed, and Virtual Reality (AR/MR/VR, or inclusively referred to as Extended Reality: XR) may become relevant in the future. We close our review with an overview of interfaces that visualize data in virtual environments. Without claiming this list to be exhaustive, we aim to provide the reader with a selective list of interesting pointers in hopes that they will ignite research topics relevant to search interfaces. Research students may find this section useful as it selectively points to a range of potential future research directions.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

6.1 Conversational Search Interfaces

Research on conversational search systems has received more attention recently partly due to the recent interest in the application of deep learning methods to NLP applications, such as chatbots and intelligent personal assistants. One of the primary objectives of conversational search systems is to provide information interactively in information-seeking conversations similar to human-human interactions. As such, user interfaces for conversational search systems are ideally similar to natural dialog interactions (Hearst, 2009) in which user's questions can be clarified during conversations. This thread of research has been pursued in IIR and has received more interest from system designers in recent years.

From the IIR perspective, researchers attempt to identify the communicative functions and purposes of elicitations (*i.e.*, questions to request information) in information-seeking conversations. For example, Wu (2005) identified the purposes of elicitations, conceptualized them as micro-level information seeking attempts, and found that user's elicitation behavior is affected by individual differences, such as status, age, and experience, and interacts with situational variables, such as interaction time and the number of utterances. Further studies (Wu and Liu, 2003; Wu and Liu, 2011) developed elicitation styles characterized by linguistic forms, utterance purposes, and communicative functions and made connections between them and user satisfaction in IR interactions. Belkin *et al.* (1993) specifically considered information-seeking strategies when designing a prototype interface in support of different kinds of user search behavior.

From the system design perspective, Radlinski and Craswell (2017) proposed system requirements for conversational search systems. Piccolo *et al.* (2019) suggested that system considerations for the design of chatbots would include tasks, interaction style, and trust. Similarly, Neururer *et al.* (2018) proposed that key features of intelligent agents like chatbots include trust, authenticity, and transparency. In a study that observed people's interactions in a laboratory setting, Trippas *et al.* (2018) compared human-human interactions with well-established search models to inform the design of spoken conversational search systems. Vtyurina *et*

The version of record is available at: http://dx.doi.org/10.1561/1500000073

al. (2017) explored the system requirements of intelligent conversational assistants for the purpose of improving user experience. Trippas *et al.* (2019) studied how intelligent assistants are used in the workplace to inform intelligent assistant design. Thomas *et al.* (2018) attempted to identify the conversational styles between users and intermediaries for the purpose of building computational models at scale for speech-based conversational agents.

As part of conversational search system interfaces, voice user interfaces have been concerned with user perception of the interface and the analysis of voice queries. While voice user interfaces are perceived as natural and convenient (Cordasco *et al.*, 2014), the speech recognition errors, user's language proficiency, and lack of interface support for voice query inputs have limited the application of voice-based systems. For example, in the investigation of voice queries for query formulation/reformulation support, Jiang *et al.* (2013)'s results indicate that query length correlates with speech recognition errors and interfaces should be designed to support voice query inputs. Guy (2018)'s analysis of voice search logs from a mobile web search engine application revealed that voice search queries are characterized by short interactions and specific queries. He further discovered that mobile user interactions lead to queries that require less interaction with the touchscreen. Kiesel *et al.* (2018) explored the relationship between voice query and user satisfaction. The findings suggest that the user's language proficiency and the number and length of answers affect user satisfaction. Myers *et al.* (2019) examined the relationship between user characteristics and user performance metrics when users interact with unfamiliar voice user interfaces.

Following the paradigm of computers as social actors, Feine *et al.* (2019) built a taxonomy of social cues of conversational agents based on interpersonal communication theories. Using ethnomethodology and conversational analysis, Porcheron *et al.* (2018) explored the role of conversational interfaces in everyday life. The findings suggest implications for request and response design of embedded social interactions within the voice user interface design.

Conversational agents have also been implemented in the healthcare domain. In a systematic review of conversational agents in health-

The version of record is available at: http://dx.doi.org/10.1561/1500000073

care, Laranjo *et al.* (2018) found that conversational agents can be characterized by task-orientation, dialog management, and dialog initiative; the findings indicate that technical performance and user experience were evaluated using a wide variety of measures. However, the efficacy and health outcomes of using these systems have rarely been evaluated. In evaluating the feasibility of using conversational assistants for medical information, Bickmore *et al.* (2018) found that the validity of medical information is critical and that system recommendations should not be unconstrained. These studies suggest that patient safety is of particular concern in the application of conversational agents or assistants in the domain of healthcare.

From the technical perspective, research on conversational search systems has focused on identifying user intent in information-seeking conversations, designing user interfaces for different interaction modes, and clarification. For example, Qu *et al.* (2019a) found that structural features (*i.e.*, the position of an utterance in a dialog) contribute the most to identifying user intent, using neural classifiers. Qu *et al.* (2019a) investigated how users perceive interfaces for presenting question-answer pairs and how they judge the quality of the answers. Braslavski *et al.* (2017) studied the generation of clarification questions from community question-answer websites by formulating the tasks as noun phrase ranking problems. Aliannejadi *et al.* (2019b) used neural models to generate clarification questions by considering sequences of purposes of interaction. Vakulenko *et al.* (2019) proposed a formal model of information seeking dialogs that consist of query, request, feedback, and answer for identifying the frequency of sequence patterns. Hashemi *et al.* (2020) investigated user responses to clarifying questions by enriching the user-system conversations with neural networks and external sources.

6.2 Searching for Leisure

Information seeking and search behaviors are historically task and goal-oriented activities that aim to resolve an information need (Wilson, 2016). While research about information interaction, including the development of models and theory, is often exclusively focused on work environments (Fisher *et al.*, 2005), there is increasing recognition of widening

The version of record is available at: http://dx.doi.org/10.1561/1500000073

this view to everyday-life information seeking (Davenport, 2010) and the more casual and leisure-oriented aspects of our lives (Stebbins, 2007; Fulton and Vondracek, 2009; Elsweiler *et al.*, 2011; Mansourian, 2019)[1]. Reasons for this low level of interest in leisure-based information behavior in the past have been identified as the stereotypical assumption that leisure may indicate unimportant, frivolous, or even banal information interaction (Stebbins, 2007). Recent technological advancements may also play an enabling role since the (mobile) web now provides access to rich and interactive media services (*e.g.*, video search, media streaming, social networks) to large crowds (Elsweiler *et al.*, 2011). In the context of the social sciences, Stebbins (2007) differentiates three categories of leisure-centered information interaction: *serious leisure*, *casual leisure*, and *project-based leisure* together with a classification of activities for these groups in the form of a conceptual framework[2]:

- *Serious leisure* activities are signified by persistence and effort (*e.g.*, becoming knowledgeable, skilled, or experienced), a career-like structure, and offer self-developing benefits, ethos, and identity. For these purposes, people use information systems to acquire knowledge, train their skills, acquire access to written and online material, and allow themselves to interact with others in the distribution of information and exchange of ideas. Mansourian (2019) further differentiates serious leisure into intellectual activities, creating and collecting, and experiences.

- *Casual leisure* includes activities that are practical, short-termed, hedonistic (and therefore focused on immediate rewards), require less knowledge, and offer immediate benefits such as play, relaxation, entertainment, sociable conversation, sensory stimulation, casual volunteering, and pleasurable aerobic activity.

- *Project-based leisure* activities are short-term, singular (or at most infrequent), and require planning and knowledge, but lack constant

[1] The reader may also refer to Stebbins' (2009) work that draws out the most important aspects from the more elaborate (Stebbins, 2007).

[2] The Serious Leisure Perspective website: https://www.seriousleisure.net

The version of record is available at: http://dx.doi.org/10.1561/1500000073

deeper commitment (*e.g.*, organizing a one-off event such as a wedding).

Research directly relating to these categories is largely concerned with human information behavior in people's natural environment, such as during gourmet cooking (Hartel, 2010), coin collecting (Case, 2010) or when consuming and enjoying music (Vinchira, 2019). In this context, studies often use ethnographic approaches (Hartel, 2010), diary studies (Wilson *et al.*, 2012b; Elsweiler *et al.*, 2011), observations, and interviews (Vinchira, 2019) to characterize the situational context of information seeking and searching for leisure activities.

These studies identify and refine fundamental frameworks and form important requirements that clarify what people need from an information system to support them during leisure-related activities. However, there is a gap between the descriptive models developed in these studies and the normative models most IR researchers rely on for system development (Fidel, 2012a). While studies certainly shed light on how such systems are used in leisure-related contexts that offer insights for improving search interfaces, they have often not been explicitly developed and evaluated for leisure-related information search.

Elsweiler *et al.* (2011) report on two studies on casual information behavior in the areas of television and social media (*i.e.*, Twitter) that are later used for developing a model for casual information behavior. Both studies show that participants focus less on information and more on experience. The information the user retrieved (*e.g.*, a movie) was often irrelevant as long as a particular experience or reaction was achieved (*e.g.*, relaxation from work or distraction from a laborious household tasks). The authors identify that participants' motivations were mostly personal (*e.g.*, learning a recipe) or contextual (*e.g.*, being in the mood for being thrilled, having time to kill, or looking for content that satisfies others). This highlights a stark difference to classic information-seeking where the information need is a driving force and a key source in its evaluation (Marchionini, 1995)[3].

Mikkonen and Vakkari (2016) analyzed casual leisure search behavior by looking at the interactions of readers of fiction with online library

[3]See also Section 2.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

catalogs. They used two existing Finnish library catalogs: the classic Satakirjaastot (Sata) catalog and the enriched BookSampo (Sampo) catalog. Sata is a traditional library catalog with basic and advanced search, customer records, and links to connected libraries. Sampo offers multiple entry points for browsing and searching books by their content (*e.g.*, browsing book covers with a carousel or via a tag cloud (see Figure 6.1)) or via the social structure of its readers (*e.g.*, browsing the virtual bookshelves of others). The authors formulated their findings from user behaviors into system guidelines for creating search interfaces for casual search. They recommend creating an appealing early starting point from which users could search and explore the catalog. Multiple access points allow readers to reach wider areas of the collection as they would otherwise be restricted to only well-known authors and titles. Combining basic and advanced search that can be extended with query terms from different content dimensions (*e.g.*, theme, emotion) provides a benefit to both objective and subjective indexing in support of such user interactions. In a similar study, Mikkonen and Vakkari (2015) investigated fiction readers' search behavior for query reformulation and browsing tasks in the same library catalogs.

Figure 6.1: BookSampo book cover carousel and a subject term cloud [picture taken from Mikkonen and Vakkari (2016)]

Hosey *et al.* (2019) explored users' information behavior with the Spotify music streaming search interface. Users searched for a variety of goals to either listen, organize, share, or check on facts. The

The version of record is available at: http://dx.doi.org/10.1561/1500000073

study evaluated users' search experiences based on success and effort. They found that the usefulness of navigational features depends on the user's current mindset or intent. A focused user (*e.g.*, looking for one particular song) determines success in binary form and users are quite intolerant of everything that reduces their effort (*e.g.*, too many clicks to perform a simple lookup-up task). Somebody who is openly exploring a music collection or who follows an artist determines success in a much wider context and is therefore willing to contribute extra effort. Using the example of 'artist search', Hosey *et al.* (2019) provide design recommendations for mobile search interfaces for a leisure-related music experience and suggest a division of the mobile screen space based on different mindsets. The top of the screen that is mostly used by focused users would benefit from a query auto-suggestion feature (as such users hardly ever use SERPs) to provide an artist link and known songs or albums from that artist. The middle section, mostly used by the open-minded user, would include playlists or album information for that artist. The bottom section, mostly used by explorative users, could offer a link to the genre used by that artist and thus invite listeners to learn more about the genre.

Wu and Hsieh (2016) investigated how people search for eBooks on a touch-wall interface (Figure 6.2). Not only did leisure searchers prefer the more relaxed clustering interface over the stricter categorization interface, they were also less anxious and more comfortable with uncertainty.

Figure 6.2: A touch-wall interface that provides utilitarian and leisure-oriented library users with access to eBooks via front covers [picture taken from Wu and Hsieh (2016)]

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Overall, these examples show that people search differently when they engage in casual leisure compared to traditional search that is focused on resolving an information need. Search interfaces for causal leisure activities need to include features that invite people to explore the information space by offering multiple and different access points, interest-based recommendations, and features for re-finding information.

Information search in the more experiential context of a museum represents one aspect of serious leisure that has been widely addressed by HCI research. Hornecker and Ciolfi (2019) overview the types of interfaces and interactions that digitally enhance users' experience in physical museums including physical installations, mobile applications, interconnected activities, and XR experiences.[4] We will revisit this category again in Section 6.4.

Sports are often the focus of serious leisure activities where people deeply engage with players, events, and particularities of matches. Sports data results are also often available to the public in remarkable detail. The visual search interface by Shao *et al.* (2016) allows users to sketch player movements in a soccer game as a search query to find matching movement patterns of previously unannotated scenes.

Citizen science, as described in Preece *et al.* (2016), is another interesting category of application where scientists and amateur volunteers have collaborated. Applications often generate data for scientific projects, such as collecting nature data as a form of serious volunteerism or casual leisure (Stebbins, 2007). To keep people engaged in these sometimes repetitive activities, user interfaces for citizen science applications need to be highly motivating. For example, Preece *et al.* (2016) invited casual visitors of a nature reserve to use a borrowed smartphone with a mobile application (see Figure 6.3) that allowed them to snap pictures that were later analyzed. While people are mainly drawn to citizen science through personal interest, the simplicity of the interface is key to keeping people engaged with the technology. This study demonstrates how a casual leisure activity — an easy-to-use mobile interface for collecting nature pictures — can be used for a serious

[4]Extended reality (XR) includes virtual reality (VR), augmented reality (AR), and mixed reality (MR).

The version of record is available at: http://dx.doi.org/10.1561/1500000073

leisure hobby in the context of citizen science. The systematic literature review by Skarlatidou *et al.* (2019) develops design guidelines of citizen science applications for scientific data collection based on an analysis of three applications: iSpot,[5] iNaturalist,[6] and Zooniverse.[7]

Figure 6.3: NatureNet mobile application that was used in the citizen science study by Preece *et al.* (2016) [picture taken from Preece *et al.* (2016)]

Similarly, Budhathoki and Haythornthwaite (2013) reviewed motivational factors in serious volunteering for OpenStreetMap, a crowdsourced effort in collecting geographic data. Their questionnaire study of over 400 contributors revealed a wide range of motivational reasons for contributing to the work including differences between serious and casual members. Serious hobbyists were driven by the community learning, the opportunity to acquire knowledge, and career-similar motivations; casual participants were drawn by the freely available data from the project.

Overall, this section mostly shows examples and studies that have investigated a) users engaging in leisure information behavior to learn about their requirements and preferences (*e.g.*, Elsweiler *et al.* (2011)) or b) the study of existing search interfaces from a leisure context

[5] https://www.ispotnature.org
[6] https://www.inaturalist.org
[7] https://www.zooniverse.org

The version of record is available at: http://dx.doi.org/10.1561/1500000073

to develop design guidelines (*e.g.*, Hosey *et al.* (2019)). In addition to this, some areas of IR also cover research that relates to leisure application contexts, such as Multimedia IR (Soleymani *et al.*, 2017) or Lifelogging (Gurrin *et al.*, 2014). Their main focus, however, is typically not on the leisure-related usage scenario that we have followed in this section. Instead, these research areas specialize in technical aspects, *e.g.*, recognising objects or people's faces in multimedia content or collecting and analysing vast amounts of personal data that can be helpful for a wide range of applications, including leisure.

6.3 Interfaces Supporting Serendipity and Creativity

There is a growing body of research about models, systems, and interfaces that allow users to more easily encounter, access, discover, and explore information in serendipitous and creative ways. Both serendipity and creativity have been considered as being interlinked in the past (Bawden, 1986).

Serendipity refers to the accidental but desirable encounter with useful and interesting resources (information, real-world objects, and people) that lead to new insights (McCay-Peet *et al.*, 2015; Björneborn, 2017). While this research relates to leisure (see Section 6.2) it is also highly relevant for work situations (McCay-Peet *et al.*, 2015). Serendipity has been linked with information behavior early on but, like searching for leisure, has only recently received wider attention. Marchionini (1995) identifies information seeking as being "both systematic and opportunistic" where "search exhibits algorithms, heuristics, and serendipity" based on strategic decisions of the information seeker during the search. Mansourian (2019) also references a wider definition of human information behavior that includes serendipity. McCay-Peet and Toms (2017) conducted a systematic review of empirical research on serendipity. Serendipity has further been linked to human information behavior through the more comprehensive model of *information encountering* that formulates users' contextual aspects as well as the process of acquiring information serendipitously (Erdelez, 1995; Erdelez and Makri, 2020).

The version of record is available at: http://dx.doi.org/10.1561/1500000073

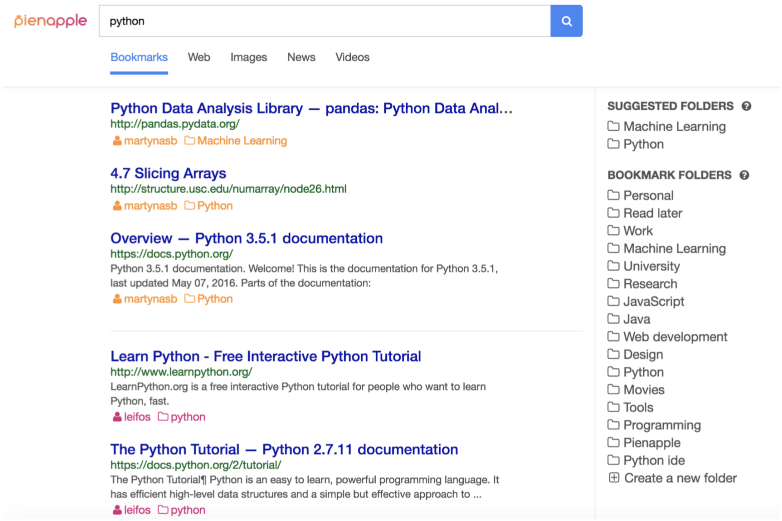


Figure 6.4: Pienapple Search interface [picture taken from Buivys and Azzopardi (2016)]

The Pienapple search interface (Figure 6.4) by Buivys and Azzopardi (2016) supports serendipity by connecting bookmark search and management with web search in a way that minimizes the user's effort for searching, bookmarking, and re-finding and allows for the collection of creative inspiration from other's bookmarks, similar to social bookmarking services. While providing a Bing search feature, it also integrates its own bookmarks and allows the user to search the public bookmarks of others.

Rahman and Wilson (2015) built a search engine that applied social media (*i.e.*, Facebook) data to highlight search results, thus connecting it with personal interests. Participants in a study were asked to exclusively use this search engine for a week. The naturalistic methodology allowed the researchers to identify queries relevant to both leisure and work-related topics. While participants with more data in their social media profiles had more serendipitous encounters, most of these took place during work-related searches.

Dörk *et al.* (2012) presented PrivotPath (Figure 6.5), an interactive visualization tool that allows people to explore highly faceted infor-

The version of record is available at: http://dx.doi.org/10.1561/1500000073

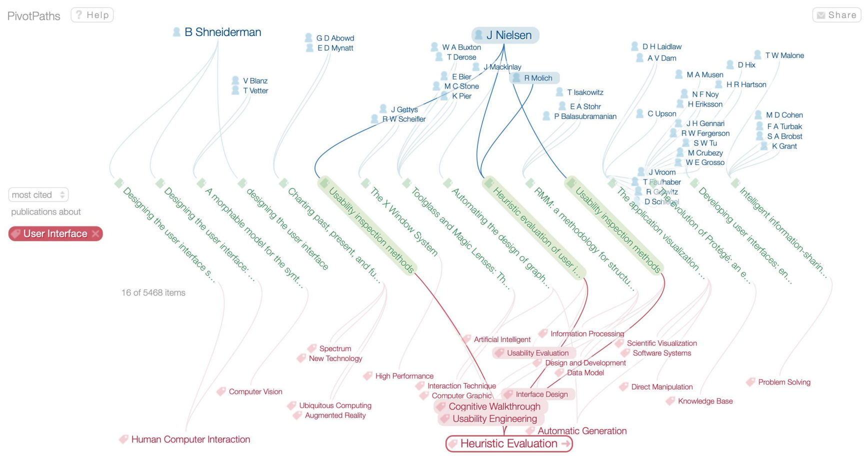


Figure 6.5: PivotPaths visualization interface supports exploration and serendipity [picture taken from Dörk *et al.* (2012)].

mation encouraging potential serendipitous encounters. Participants drawn from a research institute evaluated the tool in everyday use. They specifically enjoyed the casually interconnected and integrated style of how information was represented and started to question the current role of search services.

Taramigkou *et al.* (2017) developed an interactive exploratory search tool to support creativity by search space and result diversification, and by visualizing clues with a word cloud. Formative and summative evaluation studies were conducted in controlled laboratory and naturalistic settings to determine user preferences.

The theoretical framework developed by Björneborn (2017) explores serendipity as an affordance that exists as a relation between environmental and personal factors and informed by relevant literature. One example of empirical work on serendipity is McCay-Peet *et al.* (2015) who used a web survey to conduct a quantitative review of the causes of serendipity in relation to people and (digital) environments. While the number of serendipitous experiences depends on the type of digital environment (*e.g.*, social media, search engine), the authors further identified three qualities that such environments should offer. Information systems should provide a wide range of interesting sources,

The version of record is available at: http://dx.doi.org/10.1561/1500000073

ideas, or types of information (so-called triggers), enable connections, and actively link such information (*e.g.*, as is done in social networks), while leading to the unexpected (*e.g.*, search engines and recommender systems that deliberately diversify personalized results with alternative but potentially interesting content). In contrast, personal attributes of the participants, such as their openness to new experiences or their feeling of locus of control, had only a limited effect on their experience of serendipity (McCay-Peet *et al.*, 2015).

Frich *et al.* (2018) review HCI-based research in creativity based on ACM Digital Library publications. While research publications in this space have constantly increased since 1990, work has mostly focused on collaborative creativity, such as the workshop on digital tools in collaborative creative work (Dalsgaard *et al.*, 2018).

Zhang and Capra (2019) investigated the link between the search and creative work. Based on survey data, the authors develop a model to quantify the relationship between different creative task domains (*e.g.*, literature), resources and tools (*e.g.*, search engines), and creative process stages (*e.g.*, looking up information) to develop recommendations for creative support tools. While creative activity relates to many domains, about half of all creative activities were distributed across different devices and involved multiple creative information-seeking stages. Users preferred search engines for looking up information whereas social media was used for gathering ideas.

Lately, research in IR has discovered diversification (Santos *et al.*, 2015) and de-personalization (Bierig and Caton, 2019) as possible means by which to relax the highly engineered focus of search services. So-called "filter bubbles" (or "echo chambers") (Pariser, 2011) can lead to polarization that reduce the positive effects of serendipitous encounters and informational diversity.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

6.4 Searching in Immersive Environments

The success of Extended Reality (XR) applications[8] in the past decade has been triggered by an array of consumer head-mounted displays for VR and the ever-improving smartphone market with sensor-rich features that can readily be used for AR applications and, to some extent, for VR. Currently, there are over 4000 exclusive VR software products (mostly games)[9]. Fortune Business Insights (2019) reports that in 2018, about 40% of VR applications were in the areas of gaming and entertainment with applications for training and simulation gaining momentum. A recent survey among 650 startup founders in XR technology predicts the most disruption in the fields of healthcare (38%) and education (28%) (Perkins Coie, 2020). While applications in VR are still very much in their early stage, these projections indicate a bright future. Immersive search interfaces using XR technology are a natural consequence of this development. Currently, only limited work has been done in this area and search interfaces are mostly implemented in 2D or 3D and displayed on a screen. For example, a study on visual search used screen-based 3D visualizations that provided search functionality through different forms of interaction methods (Christmann *et al.*, 2010). Their results indicate that there is some benefit to perspective views in facilitating visual search, showing that the interaction method (manipulating a 3D object vs. immersive movement) made little difference. While 3D interfaces are well known in HCI (LaViola *et al.*, 2017), there are remarkably few search interfaces for XR technology and only a few studies that explore their benefits in this new context.

Ajanki *et al.* (2010) developed a prototype platform for extracting contextual information and providing it to users through AR devices, such as head-mounted displays or smartphones. An initial user study revealed mostly technical issues with the display, such as the readability of the text, and issues with comfort – limitations that are expected to

[8]Extended Reality (XR) combines various degrees of reality and virtuality on a spectrum and thus combines Augmented Reality (AR), Mixed Reality (MR), and Virtual Reality (VR) in a singular model (Milgram and Kishino, 1994).

[9]At the time of completing this review, the popular Steam platform offers about 4000 products that are tagged as "VR Only".

The version of record is available at: http://dx.doi.org/10.1561/1500000073

disappear in the next few years with stronger graphics cards, higher resolutions, and optimized rendering techniques.[10]

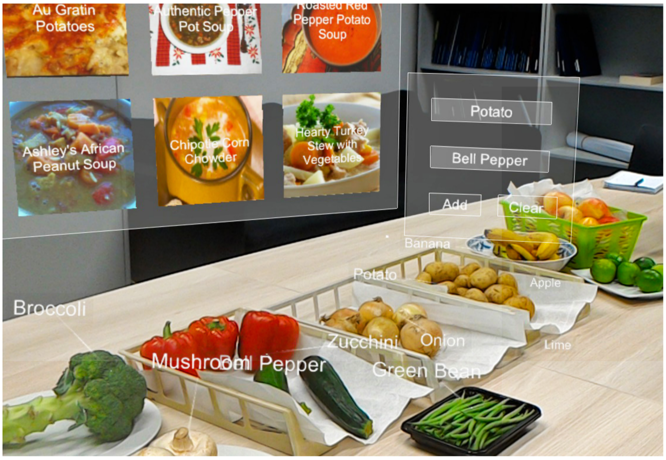


Figure 6.6: Mixed Reality search interface for a recipe search prototype that shows food tags (foreground), result lists (left), and query (right) [picture taken from Büschel *et al.* (2018)]

In Büschel *et al.* (2018), information from user query interactions and from search results overlap into a natural environment (*e.g.*, a shop, or a simulation thereof). Their work presents Mixed Reality (MR) prototypes for two different application contexts (shown in Figure 6.6) – 1) situated search for photographs (taking pictures from ones current view to search for similar images) and 2) recipe search (searching for ingredients in a shop, for cooking a meal).

Ward and Capra (2020) investigated the spatial aspects of search interfaces in VR using result displays that required various degrees of head and body movements. The 36 participants of the study were asked to find results in prepared search results. Results were either presented as a simple vertical *list* (requiring no head or body movement), a curved *grid* of 140° (requiring only head movement), and a surrounding *arch* of 220° (requiring head and body movement). Participants performed

[10]The HTC Vive Pro Eye VR headset, for example, uses eye tracking for *foveated rendering* where high-quality rendering is optimised based on where people are looking. This can help to increase screen resolutions without the need for more graphical processing power.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

equally well in all three SERP display types when all results where presented; they performed faster with the grid and the arch when results were missing. People generally preferred the list and the grid over the arch as they required less body movement.

In addition to these initial, technical investigations into XR-related search scenarios, there is also increasing interest in using XR in library contexts. Cook (2018) identifies the potential of VR for library contexts with respect to information embodiment (*e.g.*, through physical browsing of books and collection artifacts) and how it facilitates serendipity (also see Section 6.3). The paper highlights initiatives and projects at the University of Oklahoma that address how VR can provide physical and spatial features for digital content. The Oklahoma Virtual Academic Laboratory (OVAL) allows users to upload, create, and edit 3D content collaboratively in VR (Figure 6.7).

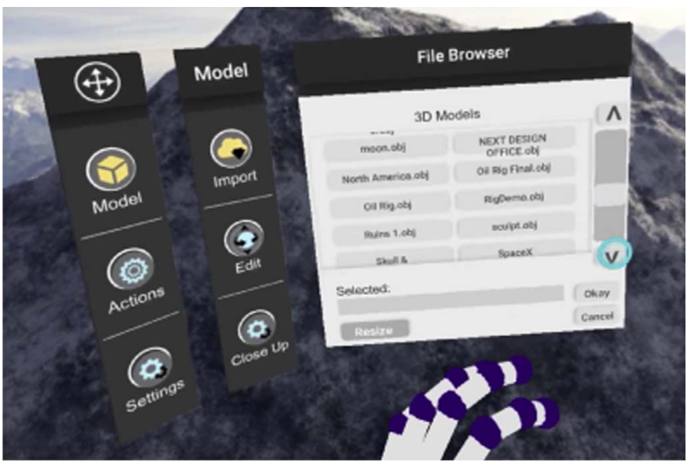


Figure 6.7: The Oklahoma Virtual Academic Laboratory (OVAL) browser allows library users to upload 3D content to make them collaboratively accessible and editable in VR [picture taken from Cook (2018)].

Museums and cultural heritage projects have applied XR technology for many years and remain active spaces for creative explorations of the medium. Hornecker and Ciolfi (2019) survey research on XR technology in museums that allow visitors to experience, interact with, and reconstruct museum artifacts and events. Common themes for the

The version of record is available at: http://dx.doi.org/10.1561/1500000073

use of XR are time travel (where visitors can explore a past-time place or event), character impersonation (where visitors experience a world or an event through a simulated character), reveal (where the visitor can dissect layers of an artifact or a place), and reconstruction (where the visitor can experience a current site (*e.g.*, the ruin of a castle) in a different state (*e.g.*, fully built and active, as it once was) (Hornecker and Ciolfi, 2019). Most recently, Marin-Morales *et al.* (2019) compared a physical museum with a virtual counterpart in terms of users' experience of presence and navigational behavior. Jerald (2015) provides general design guidelines for VR covering aspects such as perception, content creation, and interaction (and interaction design) that may be useful for designing immersive search interfaces. Overall, the use of immersive technology for search applications and services are still at an early stage with much still left to explore.

6.5 Data Visualization Interfaces in Virtual Environments

Data visualization has been identified as one of the emerging research areas in the digital humanities curricula (Walsh *et al.*, 2021). From a user perspective, comprehension of data visualization is concerned with the users' interpretation of quantitative information in the graph. The processes of graph comprehension emerge from integrated, sequential sub-processes like encoding graphical descriptions, information search, and reasoning (Carpenter and Shah, 1998; Körner *et al.*, 2014; Pinker, 2020). People's cognitive abilities, such as perceptual speed and verbal working memory are correlated with eye-tracking measures (Conati *et al.*, 2020; Toker *et al.*, 2017). Prior information-seeking experience is correlated with search performance and eye gaze behavior (Liu and Wacholder, 2017; Wittek *et al.*, 2016). Specifically, users benefit from system-driven customization of the information content presented in a data visualization system, but the degree of benefit depends upon the user characteristics of visualization literacy and locus of control (Lallé and Conati, 2019).

Data visualization interfaces that aim to support the user in understanding complex, interconnected data are commonly used in visual analytics (Keim *et al.*, 2008; Weiskopf, 2019) and e-learning (Silva *et al.*,

The version of record is available at: http://dx.doi.org/10.1561/1500000073

2019). Eye gaze data, together with search behavior data or user characteristics, can be used to infer types of search tasks (Steichen *et al.*, 2014). Of relevance to learning contexts is that fact that when task complexity increases, the attention distribution of field-dependent (holistic) users is more equally distributed than field-independent (analytic) users (Raptis *et al.*, 2017).

In the context of information seeking, research suggests that previous search experience correlates with search performance and eye gaze behavior (Liu and Wacholder, 2017; Wittek *et al.*, 2016). In a study of conversational agents for pilots (Liu *et al.*, 2020b), the user experiment with flight school students revealed that user perceptions about the usefulness of the system and its relevance to the topic are good predictors of search performance.

From user-centered design perspectives, research has focused on the usability of visualization systems, user experience, and physiological responses in VR environments. Specifically, the user-centered design principles and techniques in VR environments have been validated (Aragon and Hearst, 2005) and applied to immersive 3D environments (Gerjets *et al.*, 2018). The use of physiological signals for building up computational models that can detect cognitive load, mental stress, and emotional state for VR environments has great potential for developing user-adaptive interfaces (Gerjets *et al.*, 2018; Pettersson *et al.*, 2019; Skarbez *et al.*, 2018). More research on the effect of individual differences in cognitive processing and user perception will provide insights into user-adaptive interface design in VR environments.

6.6 Summary

This section covered the following emerging topics in search interface research:

- *Conversational* search interfaces are an interactive way to engage and support users while looking for information. We reviewed research in this increasingly popular field that merges AI and NLP to create personal assistants.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

- Specific interfaces for *leisure* search are sparse and traditionally intermingled with worktask-related search applications. We highlighted different layers of human leisure activity and studies that demonstrate the need for specific support, together with interface examples that have been optimised for leisure search application (*e.g.*, from hobbies, sports, and citizen science).

- We reviewed prior work on the concepts of *serendipity and creativity*. We showed examples of interfaces that allow users to search and explore information to allow for unintended discovery, (*e.g.*, by diversifying search, or by exploring information facets dynamically).

- *Augmented, mixed, and virtual reality (XR)* interfaces are promising access platforms for search and information seeking activity as well as for the visualization of data as a product of search and exploration. While only initial work has as yet been done in this area, this topic has great potential for the future.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

7

Summary

This review of what has been done in search interface design and evaluation in the last 10 years is structured into six sections. First, we briefly reviewed the history of search interfaces from the perspective of user interaction. We show that at the early stage of any search interface, designers consider users' need, preferences, cognitive load, and personalization and incorporate these into the design of their search interface.

Secondly, we reviewed search behavioral models and the application of search tasks and theoretical models into search interface design. We showed that there have been some disconnects between search interface design and the application of user models. However, we also found different types of theories that serve a different role in the design of the search interface. Some user models, such as Belkin and Cool (2002)'s ISSs and Bates (1979)'s search tactics, could serve as evaluation frameworks; there have been a few studies that applied Kuhlthau's ISP model to design search interfaces that could support different search stages; information foraging theory could explain users' search behaviors on search interfaces. Besides search behavior models, there have been studies that, through the exploration and application of models from

The version of record is available at: http://dx.doi.org/10.1561/1500000073

behavioral economics and cognitive psychology, proposed evaluation measures that could better reflect users' bias during search evaluation (Liu and Han, 2020; Azzopardi, 2021). There is a call for more theoretical discussion and exploration in search behavior models that could inform the design of the search interface.

Thirdly, we reviewed studies on search interface designs by adopting Marchionini's information-seeking process model, and classified search interface features into four subprocesses: understanding and planning, searching and execution, evaluation, and use. Our review has demonstrated that there have not been enough features supporting information use, (*e.g.*, helping searchers extract information, take notes, or finish the whole task). The research in Search as Learning started to explore related issues, and several reviews also call for related research to support task completion and not only searching (White, 2016; Shah and White, 2021).

Section 4 reviews research on adaptive user interfaces, personalization, and contextualization by signal sources that include such aspects as user search behaviors, physiological signals, and various contexts. It is also necessary to design search interfaces for various groups of users, (*e.g.*, children, older adults, and people with disabilities).

Section 5 reviews research about the evaluation of search interfaces to help the reader understand when and how to select appropriate evaluation approaches and measures according to the evaluation stage and objectives. It has been suggested that there should be different evaluation objects; the researcher must determine whether to evaluate the search process or the search outcomes, as well as which aspects of the usage of search interfaces ought to be evaluated.

Section 6 provides a discussion on emerging topics in search interface research including conversational search, specific interfaces for leisure search, interfaces that allow users to search and explore information all the while allowing for serendipity and creativity, and interfaces that take advantage of immersive XR technology.

To summarize, search will be ubiquitous in the future of the emerging new technologies (*e.g.*, AI, big data, cloud computing). Search systems should be able to support various types of information acquisition, either active search for task completion, search for leisure, or search

The version of record is available at: http://dx.doi.org/10.1561/1500000073

for unintended discovery. The future of search is bright. We need to find ways to drive more research to this area through exploration and creativity in both theory and practice.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

References

Ahn, J. and P. Brusilovsky. (2013). “Adaptive visualization for exploratory information retrieval”. *Information Processing and Management*. 49(5): 1139–1164. DOI: 10.1016/j.ipm.2013.01.007.

Ahn, J., P. Brusilovsky, D. He, J. Grady, and Q. Li. (2008). “Personalized Web Exploration with Task Models”. In: *Proceedings of the 17th International Conference on World Wide Web*. *WWW ’08*. New York, NY, USA: ACM. 1–10. DOI: 10.1145/1367497.1367499.

Ajanki, A., M. Billinghurst, T. Järvenpää, M. Kandemir, S. Kaski, M. Koskela, M. Kurimo, J. Laaksonen, K. Puolamäki, T. Ruokolainen, and T. Tossavainen. (2010). “Contextual information access with augmented reality”. In: *Proceedings of the 2010 IEEE International Workshop on Machine Learning for Signal Processing, MLSP 2010*. 95–100. DOI: 10.1109/MLSP.2010.5589228.

ALA. (2016). “Information Literacy Competency Standards for Higher Education”. URL: https://alair.ala.org/handle/11213/7668#f1.

Aliannejadi, M., M. Harvey, L. Costa, M. Pointon, and F. Crestani. (2019a). “Understanding Mobile Search Task Relevance and User Behaviour in Context”. In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. *CHIIR ’19*. New York, NY, USA: ACM. 143–151. DOI: 10.1145/3295750.3298923.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Aliannejadi, M., H. Zamani, F. Crestani, and W. B. Croft. (2019b). "Asking Clarifying Questions in Open-Domain Information-Seeking Conversations". In: *Proceedings of the 42Nd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR'19*. New York, NY, USA: ACM. 475–484. DOI: 10.1145/3331184.3331265.

Allan, J., B. Croft, A. Moffat, and M. Sanderson. (2012). "Frontiers, challenges, and opportunities for information retrieval: Report from SWIRL 2012 the second strategic workshop on information retrieval in Lorne". In: *ACM SIGIR Forum*. Vol. 46. No. 1. ACM. 2–32.

Anderson, L. W., D. R. Krathwohl, P. Airasian, K. Cruikshank, R. Mayer, P. Pintrich, J. Raths, and M. Wittrock. (2001). *A taxonomy for learning, teaching, and assessing: A revision ofBloom's taxonomy of educational objectives, complete edition*. New York, NY, USA: Longman.

Anick, P. and R. G. Kantamneni. (2008). "A longitudinal study of real-time search assistance adoption". In: *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '08*. New York, New York, USA: ACM Press. 701. DOI: 10.1145/1390334.1390459.

Aragon, C. R. and M. A. Hearst. (2005). "Improving Aviation Safety with information Visualization : A Flight Simulation Study". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '05*. New York, NY, USA: ACM. 441–450. DOI: 10.1145/1054972.1055033.

Arvola, P., J. Vainio, M. Junkkari, and J. Kekäläinen. (2012). "Model for Simulating Result Document Browsing in Focused Retrieval". In: *Proceedings of the 4th Information Interaction in Context Symposium. IIIX '12*. New York, NY, USA: ACM. 238–241. DOI: 10.1145/2362724.2362764.

Athukorala, K., D. Głowacka, G. Jacucci, A. Oulasvirta, and J. Vreeken. (2016). "Is exploratory search different? A comparison of information search behavior for exploratory and lookup tasks". *Journal of the Association for Information Science and Technology*. 67(11): 2635–2651.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Aula, A. (2004). “Enhancing the readability of search result summaries”. In: *Proceedings*. Vol. 2. 1–4.

Avula, S., J. Arguello, R. Capra, J. Dodson, Y. Huang, and F. Radlinski. (2019). “Embedding Search into a Conversational Platform to Support Collaborative Search”. In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval. CHIIR ’19*. New York, NY, USA: Association for Computing Machinery. 15–23. DOI: 10.1145/3295750.3298928.

Azpiazu, I. M., N. Dragovic, M. S. Pera, and J. A. Fails. (2017). “Online searching and learning: YUM and other search tools for children and teachers”. *Information Retrieval Journal*. 20(5): 524–545. DOI: 10.1007/s10791-017-9310-1.

Azzopardi, L. (2021). “Cognitive Biases in Search”. In: *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. New York, NY, USA: ACM. 27–37. DOI: 10.1145/3406522.3446023.

Baeza-Yates, R., C. Hurtado, M. Mendoza, and U. De Chile. (2004). “Query recommendation using query logs in search engines”. *Lecture notes in computer science*. 58(12): 588–596.

Bando, L. L., F. Scholer, and A. Turpin. (2010). “Constructing query-biased summaries: A comparison of human and system generated snippets”. In: *IIiX 2010 - Proceedings of the 2010 Information Interaction in Context Symposium*. 195–204. DOI: 10.1145/1840784.1840813.

Barral, O., M. J. A. A. Eugster, T. Ruotsalo, M. M. Spapé, I. Kosunen, N. Ravaja, S. Kaski, and G. Jacucci. (2015). “Exploring peripheral physiology as a predictor of perceived relevance in information retrieval”. In: *Proceedings of the International Conference on Intelligent User Interfaces (IUI ’15)*. Vol. 20. New York: ACM. 389–399. DOI: 10.1145/2678025.2701389.

Barral, O., I. Kosunen, T. Ruotsalo, M. M. Spapé, M. J. Eugster, N. Ravaja, S. Kaski, and G. Jacucci. (2016). “Extracting relevance and affect information from physiological text annotation”. *User Modelling and User-Adapted Interaction*. 26(5): 493–520. DOI: 10.1007/s11257-016-9184-8.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Bast, H. and I. Weber. (2006). "Type less, find more". In: *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '06*. 364. DOI: 10.1145/1148170.1148234.

Bates, M. J. (1989). "The design of browsing and berrypicking techniques for the online search interface". *Online Review*. 13(5): 407–424. URL: https://pages.gseis.ucla.edu/faculty/bates/berrypicking.html.

Bates, M. J. (2016). "Many paths to theory: The creative process in the information sciences". In: *Theory Development in the Information Sciences*. Ed. by D. H. Sonnenwald. Austin, TX: University of Texas Press. 21–49.

Bates, M. J. (1990). "The berry-picking search: User interface design". In: *User Interface Design*. Ed. by H. Thimbleby. Addison-Wesley.

Bates, M. J. (2002). "The cascade of interactions in the digital library interface". *Information Processing & Management*. 38(3): 381–400.

Bates, M. J. (2007). "What is browsing-really? A model drawing from behavioural science research". 12(4). URL: http://InformationR.net/ir/12-4/paper330.html.

Bates, M. J. (1979). "Information search tactics". *Journal of the American Society for Information Science*. 30(4): 205–214. DOI: 10.1002/asi.4630300406.

Bawden, D. (1986). "Information systems and the stimulation of creativity". *Journal of Information Science*. 12(5): 203–216. DOI: 10.1177/016555158601200501.

Beinema, T. C. (2017). "Developing a Web Search Interface for the Visually Impaired". *PhD thesis*. Radboud University. 89. URL: https://theses.ubn.ru.nl/bitstream/handle/123456789/5232/Beinema%2C_T.C._1.pdf?sequence=1.

Belkin, N. J. and C. A. Cool. (2002). "Classification of interactions with information". In: *Emerging frameworks and methods. Proceedings of the Fourth International Conference on Conceptions of Library and Information Science (COLIS4). Greenwood Village, CO: Libraries Unlimited*. 1–15.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Belkin, N., P. Marchetti, and C. Cool. (1993). "BRAQUE: Design of an interface to support user interaction in information retrieval". *Information Processing and Management*. 29(3): 325–344. DOI: 10.1016/0306-4573(93)90059-M.

Belkin, N. J. (1980a). "Anomalous states of knowledge as a basis for information retrieval". *Canadian journal of information science*. 5(1): 133–143.

Belkin, N. J. (1980b). "The problem of matching in information retrieval". *Theory and application of information research*.

Belkin, N. J. (1996). "Intelligent information retrieval: whose intelligence?" *ISI*. 96: 25–31.

Belkin, N. J., R. N. Oddy, and H. M. Brooks. (1982). "ASK for information retrieval: Part I. Background and theory". *Journal of documentation*. 38(2): 61–71.

Belkin, N. J. (1977). "A Concept of Information for Information Science". *PhD thesis*. London: University of London.

Belz, A. (2009). "That ' s Nice . . . What Can You Do With It ?" *Computational Linguistics*. 35(1): 111–118.

Bennett, J. L. and D. E. Walker. (1971). *Interactive bibliographic search as a challenge to interface design*. IBM Research Laboratory. 1–16.

Berget Gerd|Sandnes, F. E. (2015). "Searching Databases without Query-Building Aids: Implications for Dyslexic Users." *Information Research: An International Electronic Journal*. 20(4).

Berget, G. and A. MacFarlane. (2019). "What Is Known About the Impact of Impairments on Information Seeking and Searching?" *Journal of the Association for Information Science and Technology*. DOI: 10.1002/asi.24256.

Berget, G., A. MacFarlane, and N. Pharo. (2021). "Modelling the information seeking and searching behaviour of users with impairments: Are existing models applicable?" *Journal of Documentation*. 77(2): 381–400. DOI: 10.1108/JD-04-2020-0049.

Berget, G., F. Mulvey, and F. E. Sandnes. (2016). "Is visual content in textual search interfaces beneficial to dyslexic users?" *International Journal of Human Computer Studies*. 92-93: 17–29. DOI: 10.1016/j.ijhcs.2016.04.006.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Berget, G. and F. E. Sandnes. (2019). “Why textual search interfaces fail: a study of cognitive skills needed to construct successful queries”. *Information Research: An International Electronic Journal*. 24(1). URL: http://informationr.net/ir/24-1/paper812.html.

Bhattacharya, N. and J. Gwizdka. (2019). “Measuring Learning During Search”. In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. New York, NY, USA: ACM. 63–71. DOI: 10.1145/3295750.3298926.

Bi, T., N. Bianchi-Berthouze, A. Singh, and E. Costanza. (2019). “Understanding the Shared Experience of Runners and Spectators in Long-Distance Running Events”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI ’19*. New York, NY, USA: Association for Computing Machinery. DOI: 10.1145/3290605.3300691.

Bickmore, T. W., H. Trinh, S. Olafsson, T. K. O’Leary, R. Asadi, N. M. Rickles, and R. Cruz. (2018). “Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant”. *J Med Internet Res*. 20(9): e11510. DOI: 10.2196/11510.

Bierig, R. and S. Caton. (2019). “Special issue on de-personalisation, diversification, filter bubbles and search”. *Information Retrieval Journal*. 22(5): 419–421. DOI: 10.1007/s10791-019-09365-w.

Bilal, D. and L.-M. Huang. (2019). “Readability and word complexity of SERPs snippets and web pages on children’s search queries”. *Aslib Journal of Information Management*. 71(2): 241–259. Ed. by H. Li-Min. DOI: https://doi.org/10.1108/AJIM-05-2018-0124.

Björneborn, L. (2017). “Three key affordances for serendipity: Toward a framework connecting environmental and personal factors in serendipitous encounters”. *Journal of Documentation*. 73(5): 1053–1081. DOI: 10.1108/JD-07-2016-0097.

Borlund, P. (2013). “Interactive information retrieval: An introduction”. *Journal of Information Science Theory and Practice*. 1(3): 12–32.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Brandt, J., M. Dontcheva, M. Weskamp, and S. R. Klemmer. (2010). "Example-centric Programming: Integrating Web Search into the Development Environment". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '10*. New York, NY, USA: ACM. 513–522. DOI: 10.1145/1753326.1753402.

Brandt, J., P. J. Guo, J. Lewenstein, M. Dontcheva, and S. R. Klemmer. (2009). "Two Studies of Opportunistic Programming: Interleaving Web Foraging, Learning, and Writing Code". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '09*. New York, NY, USA: ACM. 1589–1598. DOI: 10.1145/1518701.1518944.

Braslavski, P., D. Savenkov, E. Agichtein, and A. Dubatovka. (2017). "What Do You Mean Exactly?: Analyzing Clarification Questions in CQA". In: *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval. CHIIR '17*. New York, NY, USA: ACM. 345–348. DOI: 10.1145/3020165.3022149.

Broughton, V. V. (2017). "Faceted classification as the basis of all information retrieval". In: *Faceted classification today: theory, technology and end users: proceedings of the International UDC Seminar*. London, UK: Ergon Verlag. 149–162. URL: http://discovery.ucl.ac.uk/10038742/3/Broughton_final_Faceted%20classification.pdf.

Brown, E. (2008). "Thumbtack: A bookmarking service from Microsoft".

Bruckman, A., A. Bandlow, J. Dimond, and A. Forte. (2012). "Human-Computer Interaction for Kids". In: *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Third Edition*. Ed. by A. Sears and J. Jacko. 3rd Ed. CRC Press. Chap. 36. 841–861.

Buchanan, G. and F. Loizides. (2007). "Investigating Document Triage on Paper and Electronic Media". In: *Research and Advanced Technology for Digital Libraries*. Ed. by L. Kovács, N. Fuhr, and C. Meghini. Berlin, Heidelberg: Springer Berlin Heidelberg. 416–427.

Budhathoki, N. R. and C. Haythornthwaite. (2013). "Motivation for Open Collaboration". *American Behavioral Scientist*. 57(5): 548–575. DOI: 10.1177/0002764212469364.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Budzik, J. and K. J. Hammond. (2000). "User Interactions with Everyday Applications As Context for Just-in-time Information Access". In: *Proceedings of the 5th International Conference on Intelligent User Interfaces. IUI '00*. New York, NY, USA: ACM. 44–51. DOI: 10.1145/325737.325776.

Buettner, R., S. Sauer, C. Maier, and A. Eckhardt. (2018). "Real-time Prediction of User Performance based on Pupillary Assessment via Eye-Tracking". *AIS Transactions on Human-Computer Interaction*. 10(1): 26–60. DOI: 10.17705/1thci.00103.

Buivys, M. and L. Azzopardi. (2016). "Pienapple search: an integrated search interface to support finding, refinding and sharing". *Proceedings of the Association for Information Science and Technology*. 53(1): 1–5. DOI: 10.1002/pra2.2016.14505301122.

Büschel, W., A. Mitschick, and R. Dachselt. (2018). "Here and now: Reality-based information retrieval". In: *CHIIR 2018 - Proceedings of the 2018 Conference on Human Information Interaction and Retrieval*. Vol. 2018-March. New York, New York, USA: Association for Computing Machinery, Inc. 171–180. DOI: 10.1145/3176349.3176384.

Byström, K. and P. Hansen. (2005). "Conceptual framework for tasks in information studies". *Journal of the American Society for Information science and Technology*. 56(10): 1050–1061.

Cai, F., S. Liang, and M. de Rijke. (2014). "Time-sensitive Personalized Query Auto-Completion". In: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management - CIKM '14*. 1599–1608. DOI: 10.1145/2661829.2661921.

Cai, F. and M. de Rijke. (2016). "Learning from homologous queries and semantically related terms for query auto completion". *Information Processing & Management*. 52(4): 628–643.

Capra, R., J. Arguello, H. O'Brien, Y. Li, and B. Choi. (2018). "The Effects of Manipulating Task Determinability on Search Behaviors and Outcomes". In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. New York, NY, USA: ACM. 445–454. DOI: 10.1145/3209978.3210047.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Card, S. K., P. Pirolli, M. Van Der Wege, J. B. Morrison, R. W. Reeder, P. K. Schraedley, and J. Boshart. (2001). "Information scent as a driver of Web behavior graphs: results of a protocol analysis method for Web usability". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. 498–505. DOI: 10.1145/365024.365331.

Carpenter, P. A. and P. Shah. (1998). "A model of the perceptual and conceptual processes in graph comprehension". *Journal of Experimental Psychology: Applied*. 4(2): 75–100. DOI: 10.1037/1076-898X.4.2.75.

Case, D. O. (2010). "A Model of the Information Seeking and Decision Making of Online Coin Buyers". *Information Research*. 15(4). URL: http://informationr.net/ir/15-4/paper448.html.

Chang, S.-F., W. Chen, H. J. Meng, H. Sundaram, and D. Zhong. (1998). "A fully automated content-based video search engine supporting spatiotemporal queries". *IEEE transactions on circuits and systems for video technology*. 8(5): 602–615.

Chao, P. Y. and C. C. Lin. (2015). "Young children's storybook searching with a visualized search interface". *Electronic Library*. 33(4): 610–624. DOI: 10.1108/EL-12-2013-0219.

Chen, Y., Y. Liu, M. Zhang, and S. Ma. (2017). "User Satisfaction Prediction with Mouse Movement Information in Heterogeneous Search Environment". *IEEE Transactions on Knowledge and Data Engineering*. 29(11): 2470–2483.

Chen, Y., Y. Liu, K. Zhou, M. Wang, M. Zhang, and S. Ma. (2015). "Does vertical bring more satisfaction?: Predicting search satisfaction in a heterogeneous environment". In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM. 1581–1590.

Cheng, A., V. Raghavaraju, J. Kanugo, Y. P. Handrianto, and Y. Shang. (2018). "Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes". In: *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE. 1–5. DOI: 10.1109/CCNC.2018.8319283.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Christmann, O., N. Carbonell, and S. Richir. (2010). "Visual search in dynamic 3D visualisations of unstructured picture collections". *Interacting with Computers*. 22(5): 399–416. DOI: 10.1016/j.intcom.2010.02.005.

Clark, S. (2015). "Vector Space Models of Lexical Meaning". In: *The Handbook of Contemporary Semantic Theory*. 493–522. DOI: 10.1002/9781118882139.ch16.

Collins-Thompson, K., S. Y. Rieh, C. C. Haynes, and R. Syed. (2016). "Assessing Learning Outcomes in Web Search". In: *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*. New York, NY, USA: ACM. 163–172. DOI: 10.1145/2854946.2854972.

Common Sense Media. (2020). "The Common Sense Census: Media Use by Kids Age Zero to Eight, 2020". *Tech. rep.* San Francisco, CA, USA. URL: https://www.commonsensemedia.org/research/the-common-sense-census-media-use-by-kids-age-zero-to-eight-2020.

Conati, C., S. Lallé, M. A. Rahman, and D. Toker. (2020). "Comparing and Combining Interaction Data and Eye-tracking Data for the Real-time Prediction of User Cognitive Abilities in Visualization Tasks". *ACM Transactions on Interactive Intelligent Systems*. 10(2): 1–41. DOI: 10.1145/3301400.

Cook, M. (2018). "Virtual Serendipity: Preserving Embodied Browsing Activity in the 21st Century Research Library". *Journal of Academic Librarianship*. 44(1): 145–149. DOI: 10.1016/j.acalib.2017.09.003.

Cool, C. and N. J. Belkin. (2011). "Interactive information retrieval: History and background". In: *Interactive information seeking, behaviour and retrieval*. Ed. by I. Ruthven and D. Kelly. London: Facet. 1–14.

Cordasco, G., M. Esposito, F. Masucci, M. T. Riviello, A. Esposito, G. Chollet, S. Schlögl, P. Milhorat, and G. Pelosi. (2014). "Assessing Voice User Interfaces: The vassist system prototype". In: *2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom)*. 91–96. DOI: 10.1109/CogInfoCom.2014.7020425.

Creswell, J. W. and D. J. Creswell. (2018). *Research Design: Qualitative, Quantitative \& Mixed Methods Approaches*. Ed. by J. D. Creswell. 5th ed. Los Angeles, CA: Sage.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Cutrell, E. and Z. Guan. (2007). “What are you looking for? An eye-tracking study of information usage in web search”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 407–416.

Czaja, S. J., W. R. Boot, and W. A. Rogers. (2019). *Designing for Older Adults: Principles and Creative Human Factors Approaches*. 3rd. CRC Press. 296. URL: https://www.routledge.com/Designing-for-Older-Adults-Principles-and-Creative-Human-Factors-Approaches/Czaja-Boot-Charness-Rogers/p/book/9781138053663.

Czaja, S. J. and C. C. Lee. (2012). “Older Adults and Information Technology”. In: *The Human–Computer Interaction Handbook*. Ed. by J. A. Jacko. CRC Press. Chap. 35. 825–840. DOI: 10.1201/b11963-ch-35.

Dalsgaard, P., C. Remy, J. F. Pedersen, L. M. Vermeulen, and M. M. Biskjaer. (2018). “Digital Tools in Collaborative Creative Work”. In: *Proceedings of the 10th Nordic Conference on Human-Computer Interaction. NordiCHI ’18*. New York, NY, USA: ACM. 964–967. DOI: 10.1145/3240167.3240262.

Davenport, E. (2010). “Confessional methods and everyday life information seeking”. *Annual Review of Information Science and Technology*. 44: 533–562.

Dillon, A. (2016). “Theory for design: The case of reading”. In: *Theory development in the information sciences*. Ed. by D. H. Sonnenwald. Austin, TX: University of Texas Press. 222–238.

Dinan, E., V. Logacheva, V. Malykh, A. Miller, K. Shuster, J. Urbanek, D. Kiela, A. Szlam, I. Serban, R. Lowe, S. Prabhumoye, A. W. Black, A. Rudnicky, J. Williams, J. Pineau, M. Burtsev, and J. Weston. (2020). “The Second Conversational Intelligence Challenge (ConvAI2)”. In: *The NeurIPS ’18 Competition: From Machine Learning to Intelligent Conversations*. Ed. by S. Escalera and R. Herbrich. Cham: Springer International Publishing. 187–208.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Dodd, C., R. Athauda, and M. T. P. Adam. (2017). “Designing User Interfaces for the Elderly: A Systematic Literature Review”. In: *Australasian Conference on Information Systems*. 1–11. URL: https://www.acis2017.org/wp-content/uploads/2017/11/ACIS2017_paper_146_FULL.pdf.

Don, A., E. Zheleva, M. Gregory, S. Tarkan, L. Auvil, T. Clement, B. Shneiderman, and C. Plaisant. (2007). “Discovering interesting usage patterns in text collections”. In: *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management - CIKM ’07*. New York, New York, USA: ACM Press. 213–222. DOI: 10.1145/1321440.1321473.

Dörk, M., N. H. Riche, G. Ramos, and S. Dumais. (2012). “PivotPaths: Strolling through Faceted Information Spaces”. *IEEE Transactions on Visualization and Computer Graphics*. 18(12): 2709–2718. DOI: 10.1109/TVCG.2012.252.

Dörk, M., B. Müller, J.-E. Stange, J. Herseni, and K. Dittrich. (2021). “Co-Designing Visualizations for Information Seeking and Knowledge Management”. *Open Information Science*. 4(1): 217–235. DOI: 10.1515/opis-2020-0102.

Downs, B., A. Shukla, M. Krentz, M. S. Pera, K. L. Wright, C. Kennington, and J. Fails. (2020). “Guiding the selection of child spellchecker suggestions using audio and visual cues”. In: *Proceedings of the Interaction Design and Children Conference*. New York, NY, USA: ACM. 398–408. DOI: 10.1145/3392063.3394390.

Druin, A. (2005). “What Children Can Teach Us: Developing Digital Libraries for Children with Children”. *The Library Quarterly*. 75(1): 20–41. DOI: 10.1086/428691.

Edwards, A. and D. Kelly. (2017). “Engaged or Frustrated?: Disambiguating Engagement and Frustration in Search”. In: *Proceedings of the ACM SIGIR Conference (SIGIR ’17)*. 125–134. DOI: 10.1145/2964797.2964813.

Ehlers, J., C. Strauch, and A. Huckauf. (2018). “A view to a click: Pupil size changes as input command in eyes-only human-computer interaction”. *International Journal of Human-Computer Studies*. 119: 28–34. DOI: https://doi.org/10.1016/j.ijhcs.2018.06.003.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Eickhoff, C. (2018). "Cognitive Biases in Crowdsourcing". In: *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. WSDM '18*. New York, NY, USA: Association for Computing Machinery. 162–170. DOI: 10.1145/3159652.3159654.

Elbeshausen, S., T. Mandl, and C. Womser-Hacker. (2015). "Collaborative Information Seeking in the Context of Leisure and Work Task Situations: A Comparison of Three Empirical Studies". In: *Collaborative Information Seeking: Best Practices, New Domains and New Thoughts*. Ed. by P. Hansen, C. Shah, and C.-P. Klas. Cham: Springer International Publishing. 73–98. DOI: 10.1007/978-3-319-18988-8{_}5.

Ellis, D. (1989). "A behavioural model for information retrieval system design". *Journal of information science*. 15(4-5): 237–247.

Elsweiler, D., M. L. Wilson, and B. Kirkegaard Lunn. (2011). "Understanding Casual-Leisure Information Behaviour". In: *New Directions in Information Behaviour*. Ed. by J. H. Amanda Spink. Emerald Group Publishing Ltd. Chap. 9. 211–241. DOI: 10.1108/s1876-0562(2011)002011a012.

Erdelez, S. (1995). "Information encountering: An exploration beyond information seeking". *PhD thesis*. Syracuse University. URL: https://surface.syr.edu/it_etd/38.

Erdelez, S. and S. Makri. (2020). "Information encountering re-encountered: A conceptual re-examination of serendipity in the context of information acquisition". *Journal of Documentation*. 76(3): 731–751. DOI: 10.1108/JD-08-2019-0151.

Eugster, M. J., T. Ruotsalo, M. M. Spapé, I. Kosunen, O. Barral, N. Ravaja, G. Jacucci, and S. Kaski. (2014). "Predicting term-relevance from brain signals". In: *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '14*. New York, NY, USA: ACM. 425–434. DOI: 10.1145/2600428.2609594.

Fails, J. A., M. L. Guha, and A. Druin. (2013). "Methods and Techniques for Involving Children in the Design of New Technology for Children". *Foundations and Trends in Human-Computer Interaction*. 6(2): 85–166. DOI: 10.1561/1100000018.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Feild, H., J. Allan, and R. Jones. (2010). "Predicting searcher frustration". In: *SIGIR 2010 Proceedings - 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* New York, New York, USA: ACM Press. 34–41. DOI: 10.1145/1835449.1835458.

Feine, J., U. Gnewuch, S. Morana, and A. Maedche. (2019). "A Taxonomy of Social Cues for Conversational Agents". *International Journal of Human-Computer Studies*. 132(Dec.): 138–161. DOI: 10.1016/j.ijhcs.2019.07.009.

Fidel, R. (2012a). "Human Information Behavior and Information Retrieval: Is Collaboration Possible?" In: *Human Information Interaction: An Ecological Approach to Information Behavior*. Cambridge, MA: The MIT Press. Chap. 10. 199–224.

Fidel, R. (2012b). *Human information interaction: An ecological approach to information behavior*. MIT Press.

Fisher, K. E., S. Erdelez, and L. Mckechnie. (2005). *Theories of Information Behavior (Asist Monograph)*. Information Today, Inc.

Flickner, M., H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, *et al.* (1995). "Query by image and video content: The QBIC system". *computer*. 28(9): 23–32.

Foley, C. and A. F. Smeaton. (2010). "Division of labour and sharing of knowledge for synchronous collaborative information retrieval". *Information Processing and Management*. 46(6): 762–772. DOI: 10.1016/j.ipm.2009.10.010.

Fortune Business Insights. (2019). "Virtual Reality Market Size, Share, Growth - VR Industry Report 2026". *Tech. rep.* Fortune Business Insights. 160. URL: https://www.fortunebusinessinsights.com/industry-reports/virtual-reality-market-101378.

Foster, J. (2007). "Collaborative information seeking and retrieval". *Annual Review of Information Science and Technology*. 40(1): 329–356. DOI: 10.1002/aris.1440400115.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Fourney, A. (2015). “Web Search, Web Tutorials & Software Applications: Characterizing and Supporting the Coordinated Use of Online Resources for Performing Work in Feature-Rich Software”. *PhD thesis*. University of Waterloo, CA. URL: http://hdl.handle.net/10012/9502.

Fourney, A., M. R. Morris, and R. W. White. (2017). “Web Search As a Linguistic Tool”. In: *Proceedings of the 26th International Conference on World Wide Web. WWW '17*. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee. 549–557. DOI: 10.1145/3038912.3052651.

Freund, L. (2008). “Exploiting task-document relations in support of information retrieval in the workplace”. *PhD thesis*. University of Toronto.

Freund, L. and B. M. Wildemuth. (2014). “Documenting and studying the use of assigned search tasks: RepAST”. *Proceedings of the American Society for Information Science and Technology*. 51(1): 1–4.

Frich, J., M. Mose Biskjaer, and P. Dalsgaard. (2018). “Twenty Years of Creativity Research in Human-Computer Interaction: Current State and Future Directions”. In: *Proceedings of the 2018 Designing Interactive Systems Conference. DIS '18*. New York, NY, USA: ACM. 1235–1257. DOI: 10.1145/3196709.3196732.

Fu, W. T. (2020). “How Cognitive Computational Models Can Improve Information Search”. In: *Understanding and Improving Information Search: A Cognitive Approach*. Ed. by W. T. Fu and H. van Oostendorp. Cham: Springer International Publishing. 29–45. DOI: 10.1007/978-3-030-38825-6{_}3.

Fulton, C. and R. Vondracek. (2009). “Introduction: Pleasurable pursuits: Leisure and lis research”. *Library Trends*. 57(4): 611–617. DOI: 10.1353/lib.0.0061.

Gaikwad, M. and O. Hoeber. (2019). “An Interactive Image Retrieval Approach to Searching for Images on Social Media”. In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. ACM. 173–181. DOI: 10.1145/3295750.3298930.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Gajos, K. Z., D. S. Weld, and J. O. Wobbrock. (2010). "Automatically generating personalized user interfaces with Supple". *Artificial Intelligence*. 174(12-13): 910–950. DOI: 10.1016/j.artint.2010.05.005.

Gerjets, P., M. Lachmair, M. V. Butz, and J. Lohmann. (2018). "Knowledge Spaces in VR: Intuitive Interfacing with a Multiperspective Hypermedia Environment". In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 555–556. DOI: 10.1109/VR.2018.8446137.

Given, L. M. and R. Willson. (2015). "Collaboration, Information Seeking, and Technology Use: A Critical Examination of Humanities Scholars' Research Practices". In: *Collaborative Information Seeking: Best Practices, New Domains and New Thoughts*. Ed. by P. Hansen, C. Shah, and C.-P. Klas. Cham: Springer International Publishing. 139–164. DOI: 10.1007/978-3-319-18988-8{_}8.

Golovchinsky, G., M. N. Price, and B. N. Schilit. (1999). "From Reading to Retrieval: Freeform Ink Annotations As Queries". In: *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '99*. New York, NY, USA: ACM. 19–25. DOI: 10.1145/312624.312637.

Gonçalves, V. P. V. P., V. P. de Almeida Neris, S. Seraphini, T. C. M. Dias, G. Pessin, T. Johnson, and J. Ueyama. (2017). "Providing adaptive smartphone interfaces targeted at elderly people: an approach that takes into account diversity among the elderly". *Universal Access in the Information Society*. 16(1): 129–149. DOI: 10.1007/s10209-015-0429-9.

González-Ibáñez, R. (2015). "Affective Dimension in Collaborative Information Seeking". In: *Collaborative Information Seeking: Best Practices, New Domains and New Thoughts*. Ed. by P. Hansen, C. Shah, and C.-P. Klas. Cham: Springer International Publishing. 193–208. DOI: 10.1007/978-3-319-18988-8{_}10.

González-Ibáñez, R., A. Esparza-Villamán, J. C. Vargas-Godoy, and C. Shah. (2019). "A comparison of unimodal and multimodal models for implicit detection of relevance in interactive IR". *Journal of the Association for Information Science and Technology*. 0(0): 1–13. DOI: 10.1002/asi.24202.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Gooda Sahib, N., A. Tombros, and T. Stockman. (2015). "Evaluating a search interface for visually impaired searchers". *Journal of the Association for Information Science and Technology*. 66(11): 2235–2248. DOI: 10.1002/asi.23325.

"Google – Notebook FAQ". URL: https://en.wikipedia.org/wiki/Google_Notebook.

Gossen, T. (2016). "Search engines for children: Search user interfaces and information-seeking behaviour". *PhD thesis*. 1–283. DOI: 10.1007/978-3-658-12069-6.

Gossen, T. and A. Nürnberger. (2013). "Specifics of information retrieval for young users: A survey". *Information Processing and Management*. 49(4): 739–756.

Gossen, T., S. Stober, and A. Nürnberger. (2015). "Treasure Map: Search History for Young Users". In: *5th Workshop on Context-awareness in Retrieval and Recommendation (CaRR'15) in conjunction with the 37th European Conference on Information Retrieval (ECIR'15)*. Vienna.

Gurrin, C., A. F. Smeaton, and A. R. Doherty. (2014). "LifeLogging: Personal Big Data". *Foundations and Trends® in Information Retrieval*. 8(1): 1–125. DOI: 10.1561/1500000033.

Gutwin, C., A. Cockburn, and N. Gough. (2017). "A Field Experiment of Spatially-Stable Overviews for Document Navigation". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. CHI '17*. New York, NY, USA: ACM. 5905–5916. DOI: 10.1145/3025453.3025905.

Guy, I. (2018). "The Characteristics of Voice Search: Comparing Spoken with Typed-in Mobile Web Search Queries". *ACM Trans. Inf. Syst.* 36(3): 30:1–30:28. DOI: 10.1145/3182163.

Gwizdka, J. (2010). "Distribution of Cognitive Load in Web Search". *Journal of the American Society for Information Science and Technology*. 61(11): 2167–2187.

Gwizdka, J., R. Hosseini, M. Cole, and S. Wang. (2017). "Temporal dynamics of eye-tracking and EEG during reading and relevance decisions". *Journal of the Association for Information Science and Technology*. 68(10): 2299–2312. DOI: 10.1002/asi.23904.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Hagen, M., M. Potthast, M. Völske, J. Gomoll, and B. Stein. (2016). "How Writers Search: Analyzing the Search and Writing Logs of Non-fictional Essays". In: *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval. CHIIR '16*. New York, NY, USA: ACM. 193–202. DOI: 10.1145/2854946.2854969.

Häikiö, J., A. Wallin, M. Isomursu, H. Ailisto, T. Matinmikko, and T. Huomo. (2007). "Touch-based user interface for elderly users". In: *Proceedings of the 9th international conference on Human computer interaction with mobile devices and services - MobileHCI '07*. New York, New York, USA: ACM Press. 289–296. DOI: 10.1145/1377999.1378021.

Hansen, P. and K. Jarvelin. (2005). "Collaborative Information Retrieval in an information-intensive domain". *Information Processing and Management*. 41(5): 1101–1119. DOI: 10.1016/j.ipm.2004.04.016.

Hansen, P., C. Shah, and C. P. Klas. (2015). *Collaborative Information Seeking*. Ed. by P. Hansen, C. Shah, and C.-P. Klas. *Computer Supported Cooperative Work*. Cham: Springer International Publishing. 1–226. DOI: 10.1007/978-3-319-18988-8.

Harper, D. J., I. Koychev, Y. X. Sun, and I. Pirie. (2004). "Within-document retrieval: A user-centred evaluation of relevance profiling". *Information Retrieval*. 7(3-4): 265–290.

Hartel, J. (2010). "Managing documents at home for serious leisure: A case study of the hobby of gourmet cooking". *Journal of Documentation*. 66(6): 847–874. DOI: 10.1108/00220411011087841.

Harvey, M., M. Langheinrich, and G. Ward. (2016). "Remembering through lifelogging: A survey of human memory augmentation". *Pervasive and Mobile Computing*. 27(Apr.): 14–26. DOI: 10.1016/J.PMCJ.2015.12.002.

Harvey, M. and M. Pointon. (2017). "Searching on the Go: The Effects of Fragmented Attention on Mobile Web Search Tasks". In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '17*. New York, NY, USA: ACM. 155–164. DOI: 10.1145/3077136.3080770.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Hashemi, H., H. Zamani, and W. B. Croft. (2020). "Guided Transformer: Leveraging Multiple External Sources for Representation Learning in Conversational Search". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM. 1131–1140. DOI: 10.1145/3397271.3401061.

Hassan, A., R. W. White, S. T. Dumais, and Y.-M. Wang. (2014). "Struggling or exploring?: disambiguating long search sessions". In: *Proceedings of the 7th ACM international conference on Web search and data mining*. ACM. 53–62.

He, D., P. Brusilovsky, J. Ahn, J. Grady, R. Farzan, Y. Peng, Y. Yang, and M. Rogati. (2008). "An evaluation of adaptive filtering in the context of realistic task-based information exploration". *Information Processing and Management*. 44(2): 511–533.

Hearst, M. A. (1995). "TileBars". In: *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '95*. New York, New York, USA: ACM Press. 59–66. DOI: 10.1145/223904.223912.

Hearst, M. A. (2009). "'Natural' search user interfaces". *Communications of the ACM*. 54(11): 60–67. DOI: 10.1145/2018396.2018414.

Hill, H. (2013). "Disability and accessibility in the library and information science literature: A content analysis". *Library & Information Science Research*. 35(2): 137–142. DOI: 10.1016/J.LISR.2012.11.002.

Hofmann, K., B. Mitra, F. Radlinski, and M. Shokouhi. (2014). "An eye-tracking study of user interactions with Query Auto Completion". *CIKM 2014 - Proceedings of the 2014 ACM International Conference on Information and Knowledge Management*: 549–558. DOI: 10.1145/2661829.2661922.

Hornecker, E. and L. Ciolfi. (2019). "Human-Computer Interactions in Museums". *Synthesis Lectures on Human-Centered Informatics*. 12(2): i–153. DOI: 10.2200/S00901ED1V01Y201902HCI042.

Hosey, C., L. Vujović, B. St. Thomas, J. Garcia-Gathright, and J. Thom. (2019). "Just Give Me What I Want: How People Use and Evaluate Music Search". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19*. New York, NY, USA: Association for Computing Machinery. 1–12. DOI: 10.1145/3290605.3300529.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Hourcade, J. P. (2007). “Interaction design and children”. *Foundations and Trends in Human-Computer Interaction*. 1(4): 277–392. DOI: 10.1561/1100000006.

Hoven, E. den, C. Sas, S. Whittaker, E. Van den Hoven, C. Sas, and S. Whittaker. (2012). “Introduction to this Special Issue on Designing for Personal Memories: Past, Present, and Future”. *Human–Computer Interaction*. 27(1-2). DOI: 10.1080/07370024.2012.673451.

Htun, N. N., M. Halvey, and L. Baillie. (2018). “Beyond traditional collaborative search: Understanding the effect of awareness on multi-level collaborative information retrieval”. *Information Processing and Management*. 54(1): 60–87. DOI: 10.1016/j.ipm.2017.09.003.

Huurdeman, H. C. and J. Kamps. (2015). “Supporting the process: Adapting search systems to search stages”. In: *European Conference on Information Literacy*. Springer. 394–404.

Huurdeman, H. C., M. L. Wilson, and J. Kamps. (2016). “Active and Passive Utility of Search Interface Features in Different Information Seeking Task Stages”. In: *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval. CHIIR ’16*. New York, NY, USA: ACM. 3–12. DOI: 10.1145/2854946.2854957.

Huurdeman, H. C. (2017). “Dynamic compositions: Recombining search user interface features for supporting complex work tasks”. In: *CEUR Workshop Proceedings*. Vol. 1798. Technical University of Aachen. 22–25.

Huvila, I. (2010). “The Cool and Belkin faceted classification of information interactions revisited.” *Information Research: An International Electronic Journal*. 15(4): n4.

Huvila, I. (2016). “Some remarks on the possibility of extending the theory of polyrepresentation to the study of the relevance and qualities of ’things’ represented in information collections”. *Ogledi o informacijskim znanostima - Zbornik radova u č Tatjane Aparac-Jelušić (pp.*: 187–203. Ed. by L. Farkaš, D. Vican, S. F. Tanacković, and M. D. Ivanović.

Ide, E. (1967). “User Interaction with an Automated Information Retrieval System”. In: *Report No. ISR-12 to the National Science Foundation*. Vol. Section VI.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Ide, E. (1969). "Relevance Feedback in an Automatic Document Retrieval System". In: *Report No. ISR-15 to National Science Foundation.*

Ingwersen, P. (1992). *Information retrieval interaction.* Taylor Graham London. URL: https://www.univie.ac.at/frisch/isegov/aushaengUniWien/Ingwersen_IRIinteraction.pdf.

Ingwersen, P. (1994). "Polyrepresentation of information needs and semantic entities: elements of a cognitive theory for information retrieval interaction". Dublin, Ireland.

Ingwersen, P. (1996). "Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory". *Journal of documentation.* 52(1): 3–50.

Ingwersen, P. (1999). "Cognitive Information Retrieval." *Annual review of information science and technology (ARIST).* 34: 3–52.

Ingwersen, P. and K. Järvelin. (2006). *The turn: Integration of information seeking and retrieval in context.* Vol. 18. Springer Science & Business Media.

Iofciu, T., N. Craswell, and M. Shokouhi. (2009). "Evaluating the impact of snippet highlighting in search". *Understanding the User-Logging and Interpreting User Interactions in Information Search and Retrieval (UIIR-2009)*: 44.

Al-Ismail, M., M. Yamin, Y.-H. Liu, and T. Gedeon. (2019). "Learner characteristics of m-learning preferences". *International Journal of Information Technology.* 11(3): 493–505. DOI: 10.1007/s41870-019-00279-w.

Jacucci, G., O. Barral, P. Daee, M. Wenzel, B. Serim, T. Ruotsalo, P. Pluchino, J. Freeman, L. Gamberini, S. Kaski, and B. Blankertz. (2019). "Integrating neurophysiologic relevance feedback in intent modeling for information retrieval". *Journal of the Association for Information Science and Technology.* 70(9): 917–930. DOI: 10.1002/asi.24161.

Jakobsson, M. (1986). "Autocompletion in full text transaction entry: A method for humanized input". In: *Conference on Human Factors in Computing Systems - Proceedings.* New York, New York, USA: ACM Press. 327–332. DOI: 10.1145/22627.22391.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Järvelin, K., R. Savolainen, E. Sormunen, P. Vakkari, P. Arvola, F. Baskaya, A. Järvelin, J. Kekäläinen, H. Keskustalo, and S. Kumpulainen. (2015). "Task-Based Information Interaction Evaluation". *Acm Transactions on Information Systems*. 33(1): 1–30.

Jerald, J. (2015). *The VR Book*. Morgan \& Claypool. DOI: 10.1145/2792790.

Jiang, J., W. Jeng, and D. He. (2013). "How do users respond to voice input errors? Lexical and phonetic query reformulation in voice search". *SIGIR 2013 - Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*: 143–152. DOI: 10.1145/2484028.2484092.

Jiang, J.-Y., Y.-Y. Ke, P.-Y. Chien, and P.-J. Cheng. (2014). "Learning User Reformulation Behavior for Query Auto-Completion". In: *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval. SIGIR '14*. New York, NY, USA: Association for Computing Machinery. 445–454. DOI: 10.1145/2600428.2609614.

Jiang, T., Y. Chi, and H. Gao. (2017). "A clickstream data analysis of Chinese academic library OPAC users' information behavior". *Library and Information Science Research*. 39(3): 213–223. DOI: https://doi.org/10.1016/j.lisr.2017.07.004.

Johnson, J. and K. Finn. (2017). *Designing User Interfaces for an Aging Population*. Elsevier. DOI: 10.1016/c2015-0-01451-4.

Jürgens, J. J., C. Womser-Hacker, and T. Mandl. (2014). "Modeling the Interactive Patent Retrieval Process: An Adaptation of Marchionini's Information Seeking Model". In: *Proceedings of the 5th Information Interaction in Context Symposium. IIiX '14*. New York, NY, USA: Association for Computing Machinery. 247–250. DOI: 10.1145/2637002.2637034.

Kalbach, J. (2006). ""I'm feeling lucky": The role of emotions in seeking information on the Web". *Journal of the American Society for Information Science and Technology*. 57(6): 813–818. DOI: 10.1002/asi.20299.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Karaman, Ç. Ç. and T. M. Sezgin. (2018). "Gaze-based predictive user interfaces: Visualizing user intentions in the presence of uncertainty". *International Journal of Human Computer Studies*. 111: 78–91. DOI: 10.1016/j.ijhcs.2017.11.005.

Katsini, C., C. Fidas, G. E. Raptis, M. Belk, G. Samaras, and N. Avouris. (2018). "Eye Gaze-driven Prediction of Cognitive Differences during Graphical Password Composition". In: *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - IUI '18*. New York, New York, USA: ACM Press. 147–152. DOI: 10.1145/3172944.3172996.

Katter, R. V. and D. B. McCARN. (1971). "AIM-TWX: an experimental on-line bibliographic retrieval system". *Interactive bibliographic search: The user/computer interface*: 121–141.

Keim, D., G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. (2008). "Visual Analytics: Definition, Process, and Challenges". In: *Information Visualization. Lecture Notes in Computer Science, vol 4950*. Ed. by A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North. Berlin, Heidelberg: Springer Berlin Heidelberg. 154–175. DOI: 10.1007/978-3-540-70956-5{_}7.

Kellar, M., C. Watters, and M. Shepherd. (2007). "A field study characterizing web-based information-seeking tasks". *Journal of the American Society for Information Science and Technology*. 58(7): 999–1018.

Kelly, D. (2009). "Methods for evaluating interactive information retrieval systems with users". *Foundations and trends in Information Retrieval*. 3(1-2): 1–224.

Kelly, D. (2015). "Human Information Interaction & Retrieval". *Tech. rep.* URL: https://mklab.iti.gr/essir2015/wp-content/uploads/2015/03/ESSIR2015_Kelly.pdf.

Kelly, D., J. Arguello, and R. Capra. (2013). "NSF workshop on task-based information search systems". In: *ACM SIGIR Forum*. Vol. 47. No. 2. ACM. 116–127.

Kelly, D. and X. Fu. (2007). "Eliciting better information need descriptions from users of information search systems". *Information Processing and Management*. 43(1): 30–46.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Kelly, D., K. Gyllstrom, and E. W. Bailey. (2009). “A comparison of query and term suggestion features for interactive searching”. In: *Proceedings - 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009*. New York, New York, USA: ACM Press. 371–378. DOI: 10.1145/1571941.1572006.

Khajouei, R., M. Zahiri Esfahani, and Y. Jahani. (2017). “Comparison of heuristic and cognitive walkthrough usability evaluation methods for evaluating health information systems”. *Journal of the American Medical Informatics Association*. 24(e1): e55–e60. DOI: 10.1093/jamia/ocw100.

Kiesel, J., A. Bahrami, B. Stein, A. Anand, and M. Hagen. (2018). “Toward Voice Query Clarification”. *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR ’18*: 1257–1260. DOI: 10.1145/3209978.3210160.

Kim, J. (2006). “Task difficulty as a predictor and indicator of web searching interaction”. In: *CHI’06 Extended Abstracts on Human Factors in Computing Systems*. ACM. 959–964.

Kim, S. and D. Soergel. (2006). “Selecting and measuring task characteristics as independent variables”. *Proceedings of the American Society for Information Science and Technology*. 42(1): 1–16. DOI: 10.1002/meet.14504201111.

Kiseleva, J., K. Williams, A. Hassan Awadallah, A. C. Crook, I. Zitouni, and T. Anastasakos. (2016). “Predicting User Satisfaction with Intelligent Assistants”. In: *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval - SIGIR ’16*. *SIGIR ’16*. New York, New York, USA: ACM Press. 45–54. DOI: 10.1145/2911451.2911521.

Kocaballi, A. B., L. Laranjo, and E. Coiera. (2019). “Understanding and Measuring User Experience in Conversational Interfaces”. *Interacting with Computers*. May. DOI: 10.1093/iwc/iwz015.

Koenemann, J. and N. J. Belkin. (1996). “A case for interaction”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems common ground - CHI ’96*. Ed. by M. J. Tauber. New York, New York, USA: ACM Press. 205–212. DOI: 10.1145/238386.238487.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Kopak, R., L. Freund, and H. L. O'Brien. (2010). "Supporting Semantic Navigation". In: *Proceedings of the Third Symposium on Information Interaction in Context. IIiX '10*. New York, NY, USA: ACM. 359–364. DOI: 10.1145/1840784.1840839.

Körner, C., M. Höfler, B. Tröbinger, and I. D. Gilchrist. (2014). "Eye movements indicate the temporal organisation of information processing in graph comprehension". *Applied Cognitive Psychology*. 28(3): 360–373. DOI: 10.1002/acp.3006.

Koskela, M., P. Luukkonen, T. Ruotsalo, M. SjÖberg, and P. Floréen. (2018). "Proactive Information Retrieval by Capturing Search Intent from Primary Task Context". *ACM Transactions on Interactive Intelligent Systems*. 8(3): 1–25. DOI: 10.1145/3150975.

Kowalski, J., A. Jaskulska, K. Skorupska, K. Abramczuk, C. Biele, W. Kopeć, and K. Marasek. (2019). "Older Adults and Voice Interaction". In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19. CHI EA '19*. New York, New York, USA: ACM Press. 1–6. DOI: 10.1145/3290607.3312973.

Kuhlthau, C. C. (2005). "Towards collaboration between information seeking and information retrieval". *Information Research-an International Electronic Journal*. 10(2). URL: http://informationr.net/ir/10-2/paper225.html.

Kuhlthau, C. C. (1991). "Inside the search process: Information seeking from the user's perspective". *Journal of the American Society for Information Science*. 42(5): 361–371.

Kules, B. and R. Capra. (2011). "Visualizing Stages During an Exploratory Search Session". In: *Hcir 2011*. URL: https://docs.google.com/a/kent.edu/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxoY2lyd29ya3Nob3B8Z3g6M2I2ZjI1ODYxZGZlM2RjOQ. Retrieved: 2021-11-02.

Kumpulainen, S. and K. Jarvelin. (2010). "Information interaction in molecular medicine: integrated use of multiple channels". New Brunswick, New Jersey, USA. DOI: 10.1145/1840784.1840800.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

La Barbera, D., K. Roitero, G. Demartini, S. Mizzaro, and D. Spina. (2020). "Crowdsourcing Truthfulness: The Impact of Judgment Scale and Assessor Bias". In: *Advances in Information Retrieval: ECIR 2020. Lecture Notes in Computer Science, vol 12036*. Ed. by J. M. Jose, E. Yilmaz, J. Magalhães, P. Castells, N. Ferro, M. J. Silva, and F. Martins. Cham: Springer International Publishing. 207–214.

Lallé, S. and C. Conati. (2019). "The role of user differences in customization". In: *Proceedings of the 24th International Conference on Intelligent User Interfaces - IUI '19. IUI '19*. New York, New York, USA: ACM Press. 329–339. DOI: 10.1145/3301275.3302283.

Laranjo, L., A. G. Dunn, H. L. Tong, A. B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, A. Y. S. Lau, and E. Coiera. (2018). "Conversational agents in healthcare: a systematic review". *Journal of the American Medical Informatics Association*. 25(9): 1248–1258. DOI: 10.1093/jamia/ocy072.

Larsen, B., C. Lioma, and A. de Vries. (2012). "Report on TBAS 2012". *ACM SIGIR Forum*. 46(1): 71–77. DOI: 10.1145/2215676.2215684.

LaViola, J. J., E. Kruijff, R. McMahan, D. Bowman, and I. Poupyrev. (2017). *3D User Interfaces: Theory and Practice*. 2nd. Addison-Wesley. 624.

Li, Y. and N. J. Belkin. (2008). "A Faceted Approach to Conceptualizing Tasks in Information Seeking". *Information Processing & Management*. 44(6): 1822–1837.

Li, Y. and N. J. Belkin. (2010). "An exploration of the relationships between work task and interactive information search behavior". *Journal of the American Society for information Science and Technology*. 61(9): 1771–1789.

Liu, C., N. J. Belkin, and M. J. Cole. (2012). "Personalization of search results using interaction behaviors in search sessions". In: *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. ACM. 205–214.

Liu, C., J. Liu, and N. J. Belkin. (2014). "Predicting search task difficulty at different search stages". In: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM. 569–578.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Liu, C., X. Song, H. Liu, and N. J. Belkin. (2020a). "Modeling Knowledge Change Behaviors in Learning-related Tasks". In: *CEUR Workshop Proceedings*. Vol. 2699. Galway, Ireland. URL: http://ceur-ws.org/Vol-2699/paper18.pdf.

Liu, H., P. Mulholland, D. Song, V. Uren, and S. Rüger. (2010). "Applying information foraging theory to understand user interaction with content-based image retrieval". In: *Proceedings of the third symposium on Information interaction in context*. ACM. 135–144.

Liu, J. and N. J. Belkin. (2012). "Searching vs. writing: Factors affecting information use task performance". In: *Proceedings of the American Society for Information Science & Technology*. Baltimore, Maryland. 1–10.

Liu, J. and N. J. Belkin. (2014). "Multi-aspect information use task performance: The roles of topic knowledge, task structure, and task stage". *Proceedings of the American Society for Information Science and Technology*. 51(1): 1–10. DOI: 10.1002/meet.2014.14505101031.

Liu, J., N. J. Belkin, X. Zhang, and X. Yuand. (2013). "Examining users' knowledge change in the task completion process". *Information Processing and Management*. 49(5): 1058–1074. DOI: 10.1016/j.ipm.2012.08.006.

Liu, J., H. Zamir, L. Yuan, and S. K. Hastings. (2018). "Search systems and their features: What college students use to find and save information". *Library & Information Science Research*. 40(2): 118–124.

Liu, J. and F. Han. (2020). "Investigating Reference Dependence Effects on User Search Interaction and Satisfaction". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: ACM. 1141–1150. DOI: 10.1145/3397271.3401085.

Liu, Y.-H. (2010). "A meta-analysis of the effects of search experience on search performance in terms of the recall measure in controlled IR user experiments". In: *Proceedings of the Fifteenth Australasian Document Computing Symposium*. Ed. by F. Scholer, A. Trotman, and A. Turpin. Melbourne, Australia: School of Computer Science and IT, RMIT University. 105–110. URL: http://www.cs.rmit.edu.au/adcs2010/proceedings/pdf/paper%2019.pdf.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Liu, Y.-H., A. Arnold, G. Dupont, C. Kobus, and F. Lancelot. (2020b). “Evaluation of conversational agents for aerospace domain”. In: *CEUR Workshop Proceedings*. Vol. 2621. Samatan, Gers, France. URL: http://ceur-ws.org/Vol-2621/CIRCLE20_21.pdf.

Liu, Y.-H. and R. Bierig. (2014). “A review of users’ search contexts for lifelogging system design”. In: *Proceedings of the 5th Information Interaction in Context Symposium on - IIiX ’14*. New York, New York, USA: ACM Press. 271–274. DOI: 10.1145/2637002.2637040.

Liu, Y.-H., P. Thomas, M. Bacic, T. Gedeon, and X. Li. (2017). “Natural Search User Interfaces for Complex Biomedical Search: An Eye Tracking Study”. *Journal of the Australian Library and Information Association*. 66(4): 364–381. DOI: 10.1080/24750158.2017.1357915.

Liu, Y.-H. and N. Wacholder. (2017). “Evaluating the impact of MeSH (Medical Subject Headings) terms on different types of searchers”. *Information Processing and Management*. 53(4): 851–870. DOI: 10.1016/j.ipm.2017.03.004.

Loizides, F., T. Photiades, A. Mavri, and P. Zaphiris. (2014). “On Interactive Interfaces for Semi-Structured Academic Document Seeking and Relevance Decision Making”. *New Review of Information Networking*. 19(2): 67–95. DOI: 10.1080/13614576.2014.955209.

Low, T., N. Bubalo, T. Gossen, M. Kotzyba, A. Brechmann, A. Huckauf, and A. Nürnberger. (2017). “Towards Identifying User Intentions in Exploratory Search Using Gaze and Pupil Tracking”. In: *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval. CHIIR ’17*. New York, NY, USA: ACM. 273–276. DOI: 10.1145/3020165.3022131.

MacKay, B. and C. Watters. (2008). “Understanding and supporting multi-session web tasks”. *Proceedings of the American Society for Information Science and Technology*. 45(1): 1–13.

Mansourian, Y. (2019). “How Passionate People Seek and Share Various Forms of Information in Their Serious Leisure”. *Journal of the Australian Library and Information Association*. 69(1): 17–30. DOI: 10.1080/24750158.2019.1686569.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Marchionini, G. (1989). “Information-seeking strategies of novices using a full-text electronic encyclopedia”. *Journal of the American Society for Information Science*. 40(1): 54–66. DOI: 10.1002/(SICI)1097-4571(198901)40:1<54::AID-ASI6>3.0.CO;2-R.

Marchionini, G. (1995). *Information Seeking in Electronic Environments*. Cambridge University Press. DOI: 10.1017/cbo9780511626388.

Marchionini, G. (2006). “Exploratory search: from finding to understanding”. *Communications of the ACM*. 49(4): 41–46.

Marchionini, G. and R. W. White. (2009). “Information-Seeking Support Systems [Guest Editors’ Introduction]”. *Computer*. 42(3): 30–32. DOI: 10.1109/MC.2009.88.

Marin-Morales, J., J. L. Higuera-Trujillo, C. d. J. Ripoll, C. Llinares, J. Guixeres, S. Inarra, and M. Alcaniz. (2019). “Navigation Comparison between a Real and a Virtual Museum: Time-dependent Differences using a Head Mounted Display.” *Interact. Comput.* 31(2): 208–220. DOI: 10.1093/iwc/iwz018.

Marshall, S. (2002). “The Index of Cognitive Activity: measuring cognitive workload”. In: *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants*. IEEE. 7–5. DOI: 10.1109/HFPP.2002.1042860.

Al-Maskari, A. and M. Sanderson. (2011). “The effect of user characteristics on search effectiveness in information retrieval”. *Information Processing and Management*. 47(5): 719–729. DOI: 10.1016/j.ipm.2011.03.002.

McCay-Peet, L. and E. G. Toms. (2017). *Researching Serendipity in Digital Information Environments*. Vol. 9. No. 6. Morgan & Claypool Publishers. i–91. DOI: 10.2200/S00790ED1V01Y201707ICR059.

McCay-Peet, L., E. G. Toms, and E. K. Kelloway. (2015). “Examination of relationships among serendipity, the environment, and individual differences”. *Information Processing and Management*. 51(4): 391–412. DOI: 10.1016/j.ipm.2015.02.004.

Medlar, A., J. Pyykkö, and D. Glowacka. (2017). “Towards Fine-Grained Adaptation of Exploration/Exploitation in Information Retrieval”. In: *Proceedings of the 22Nd International Conference on Intelligent User Interfaces. IUI ’17*. New York, NY, USA: ACM. 623–627.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Meho, L. I. and H. R. Tibbo. (2003). “Modeling the information-seeking behavior of social scientists: Ellis’s study revisited”. *Journal of the American society for Information Science and Technology*. 54(6): 570–587.

Merčun, T., M. Žumer, and T. Aalberg. (2017). “Presenting bibliographic families using information visualization: Evaluation of FRBR-based prototype and hierarchical visualizations”. *Journal of the Association for Information Science and Technology*. 68(2): 392–411. DOI: 10.1002/asi.23659.

Mikkonen, A. and P. Vakkari. (2015). “Books’ Interest Grading and Fiction Readers’ Search Actions during Query Reformulation Intervals”. In: *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*. Vol. 2015-June. New York, New York, USA: Institute of Electrical and Electronics Engineers Inc. 27–36. DOI: 10.1145/2756406.2756922.

Mikkonen, A. and P. Vakkari. (2016). “Finding fiction: Search moves and success in two online catalogs”. *Library and Information Science Research*. 38(1): 60–68. DOI: 10.1016/j.lisr.2016.01.006.

Milgram, P. and F. Kishino. (1994). “A Taxonomy of Mixed Reality Visual Displays”. *IEICE Transactions on Information and Systems*. E77-D(12): 1321–1329.

Mitra, B., M. Shokouhi, F. Radlinski, and K. Hofmann. (2014). “On user interactions with query auto-completion”. In: *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval - SIGIR ’14*. 1055–1058. DOI: 10.1145/2600428.2609508.

Mitsui, M., J. Liu, N. J. Belkin, and C. Shah. (2017). “Predicting information seeking intentions from search behaviors”. In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM. 1121–1124.

Mitsui, M., J. Liu, and C. Shah. (2018a). “How much is too much?: Whole session vs. first query behaviors in task type prediction”. In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM. 1141–1144.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Mitsui, M., J. Liu, and C. Shah. (2018b). "The paradox of personalization: Does task prediction require individualized models?" In: *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. ACM. 277–280.

Mollashahi, E. S., M. S. Uddin, and C. Gutwin. (2018). "Improving Revisitation in Long Documents with Two-level Artificial-landmark Scrollbars". In: *Proceedings of the 2018 International Conference on Advanced Visual Interfaces. AVI '18*. New York, NY, USA: ACM. 30:1–30:9. DOI: 10.1145/3206505.3206554.

Montoya Freire, M. L., D. Potts, N. R. Dayama, A. Oulasvirta, and M. Di Francesco. (2019). "Foraging-based optimization of pervasive displays". *Pervasive and Mobile Computing*. 55(Apr.): 45–58. DOI: 10.1016/j.pmcj.2019.02.008.

Moshfeghi, Y. and J. M. Jose. (2013). "On cognition, emotion, and interaction aspects of search tasks with different search intentions". In: *WWW 2013 - Proceedings of the 22nd International Conference on World Wide Web*. New York, New York, USA: ACM Press. 931–941. DOI: 10.1145/2488388.2488469.

Moshfeghi, Y., F. Pollick, P. Triantafillou, and F. Pollick. (2019). "Towards Predicting a Realisation of an Information Need Based on Brain Signals". In: *The World Wide Web Conference. WWW '19*. ACM. New York, NY, USA: ACM. 1300–1309. DOI: 10.1145/3308558.3313671.

Murata, A. and H. Iwase. (2005). "Usability of Touch-Panel Interfaces for Older Adults". *Human Factors: The Journal of the Human Factors and Ergonomics Society*. 47(4): 767–776. DOI: 10.1518/001872005775570952.

Murcia-López, M., T. Collingwoode-Williams, W. Steptoe, R. Schwartz, T. J. Loving, and M. Slater. (2020). "Evaluating Virtual Reality Experiences Through Participant Choices". In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 747–755. DOI: 10.1109/VR46266.2020.00098.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Myers, C. M., A. Furqan, and J. Zhu. (2019). “The Impact of User Characteristics and Preferences on Performance with an Unfamiliar Voice User Interface”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI ’19*. New York, NY, USA: ACM. 47:1–47:9. DOI: 10.1145/3290605.3300277.

Na, K. (2021). “The effects of cognitive load on query reformulation: mental demand, temporal demand and frustration”. *Aslib Journal of Information Management*. 73(3): 436–453. DOI: 10.1108/AJIM-07-2020-0206.

Needleman, R. (2009). “Yahoo Search Pad: An online notebook that watches you”. URL: https://www.cnet.com/news/yahoo-search-pad-an-online-notebook-that-watches-you/.

Neururer, M., S. Schlögl, L. Brinkschulte, and A. Groth. (2018). “Perceptions on Authenticity in Chat Bots”. *Multimodal Technologies and Interaction*. 2(3): 60. DOI: 10.3390/mti2030060.

Nielsen, J. and T. K. Landauer. (1993). “Mathematical model of the finding of usability problems”. In: *Conference on Human Factors in Computing Systems - Proceedings*. New York, New York, USA: Publ by ACM. 206–213. DOI: 10.1145/169059.169166.

Niu, X. and D. Kelly. (2014). “The use of query suggestions during information search”. *Information Processing & Management*. 50(1): 218–234. DOI: 10.1016/j.ipm.2013.09.002.

O’Brien, H. L., J. Arguello, and R. Capra. (2020). “An empirical study of interest, task complexity, and search behaviour on user engagement”. *Information Processing and Management*. 57(3): 102226. DOI: 10.1016/j.ipm.2020.102226.

Oddy, R. N. (1977). “Information retrieval through man-machine dialogue”. *Journal of documentation*. 33(1): 1–14.

Oliveira, F. T. P., A. Aula, and D. M. Russell. (2009). “Discriminating the Relevance of Web Search Results with Measures of Pupil Size”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI ’09*. New York, NY, USA: ACM. 2209–2212. DOI: 10.1145/1518701.1519038.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Ong, K. (2017). “Using information foraging theory to understand search behavior in different environments”. In: *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. ACM. 411–413.

Ottley, A. (2020). “Adaptive and Personalized Visualization”. *Synthesis Lectures on Visualization*. 7(1): 1–117.

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Group. 294.

Pauchet, S., C. Letondal, J.-L. Vinot, M. Causse, M. Cousy, V. Becquet, and G. Crouzet. (2018). “GazeForm: Dynamic Gaze-adaptive Touch Surface for Eyes-free Interaction in Airliner Cockpits”. In: *Proceedings of the 2018 Designing Interactive Systems Conference. DIS ’18*. New York, NY, USA: ACM. 1193–1205. DOI: 10.1145/3196709.3196712.

Pejtersen, A. M. (1989). *The Book House. Modelling Users’ Needs for Search Strategies as a Basis for System Design*. Frankfurt 50: INT SOC KNOWLEDGE ORGANIZATION. 233–235.

Perkins Coie. (2020). “2020 Augmented and Virtual Reality Survey Results | Perkins Coie”. *Tech. rep.* Perkins Coie and XR Association. 33. URL: https://www.perkinscoie.com/images/content/2/3/v4/231654/2020-AR-VR-Survey-v3.pdf.

Pettersson, I., M. Karlsson, and F. T. Ghiurau. (2019). “Virtually the Same Experience?: Learning from User Experience Evaluation of In-vehicle Systems in VR and in the Field”. In: *Proceedings of the 2019 on Designing Interactive Systems Conference. DIS ’19*. New York, NY, USA: ACM. 463–473. DOI: 10.1145/3322276.3322288.

Piaget, J. and B. Inhelder. (1969). *The Psychology of the Child*. 2nd Ed. Basic Books.

Piccolo, L., M. Mensio, and H. Alani. (2019). “Chasing the Chatbots: Directions for Interaction and Design Research”. In: *Internet Science. INSCI 2018*. Ed. by s Bodrunova. Vol. 11551. *Lecture Notes in Computer Science*. Springer. 157–169. URL: http://oro.open.ac.uk/57382/.

Pinker, S. (2020). “A theory of graph comprehension”. In: *Artificial Intelligence and the Future of Testing*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. 87–140. DOI: 10.4324/9781315808178-10.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Piper, A. M., R. Brewer, and R. Cornejo. (2017). “Technology learning and use among older adults with late-life vision impairments”. *Universal Access in the Information Society*. 16(3): 699–711. DOI: 10.1007/s10209-016-0500-1.

Pirolli, P. and S. Card. (1999). “Information foraging.” *Psychological review*. 106(4): 643–675.

Porcheron, M., J. E. Fischer, S. Reeves, and S. Sharples. (2018). “Voice Interfaces in Everyday Life”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. *CHI ’18*. New York, NY, USA: ACM. 640:1–640:12. DOI: 10.1145/3173574.3174214.

Portet, F., M. Vacher, C. Golanski, C. Roux, and B. Meillon. (2013). “Design and Evaluation of a Smart Home Voice Interface for the Elderly: Acceptability and Objection Aspects”. *Personal Ubiquitous Comput.* 17(1): 127–144. DOI: 10.1007/s00779-011-0470-5.

Preece, J., C. Boston, T. Yeh, J. Cameron, M. L. Maher, and K. Grace. (2016). “Enticing casual nature preserve visitors into citizen science via photos”. In: *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. Vol. 26-Februar. New York, New York, USA: Association for Computing Machinery. 373–376. DOI: 10.1145/2818052.2869104.

Qu, C., L. Yang, W. B. Croft, Y. Zhang, J. R. Trippas, and M. Qiu. (2019a). “User Intent Prediction in Information-seeking Conversations”. In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. *CHIIR ’19*. New York, NY, USA: ACM. 25–33. DOI: 10.1145/3295750.3298924.

Qu, C., C. Sas, and G. Doherty. (2019b). “Exploring and Designing for Memory Impairments in Depression”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*. New York, New York, USA: ACM Press. 1–15. DOI: 10.1145/3290605.3300740.

Radlinski, F. and N. Craswell. (2017). “A Theoretical Framework for Conversational Search”. In: *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. *CHIIR ’17*. New York, NY, USA: ACM. 117–126. DOI: 10.1145/3020165.3020183.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Ragni, M. (2020). "Artificial Intelligence and High-Level Cognition". In: *A Guided Tour of Artificial Intelligence Research*. Ed. by P. Marquis, O. Papini, and H. Prade. Cham: Springer International Publishing. 457–486. DOI: 10.1007/978-3-030-06170-8{_}14.

Rahman, A. and M. L. Wilson. (2015). "Exploring Opportunities to Facilitate Serendipity in Search". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '15*. New York, New York, USA: ACM Press. 939–942. DOI: 10.1145/2766462.2767783.

Raptis, G. E., C. Katsini, M. Belk, C. Fidas, G. Samaras, and N. Avouris. (2017). "Using Eye Gaze Data and Visual Activities to Infer Human Cognitive Styles". *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization - UMAP '17*: 164–173. DOI: 10.1145/3079628.3079690.

Rieh, S. Y. and H. (Xie. (2006). "Analysis of multiple query reformulations on the web: The interactive information retrieval context". *Information Processing and Management*. 42(3): 751–768. DOI: 10.1016/j.ipm.2005.05.005.

Rose, D. E., D. Orr, and R. G. P. Kantamneni. (2007). "Summary attributes and perceived search quality". In: *Proceedings of the 16th international conference on World Wide Web*. 1201–1202.

Roy, N., M. V. Torre, U. Gadiraju, D. Maxwell, and C. Hauff. (2021). "Note the Highlight: Incorporating Active Reading Tools in a Search as Learning Environment". In: *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. New York, NY, USA: ACM. 229–238. DOI: 10.1145/3406522.3446025.

Ruotsalo, T., J. Peltonen, M. J. A. Eugster, D. Głowacka, P. Floréen, P. Myllymäki, G. Jacucci, and S. Kaski. (2018). "Interactive Intent Modeling for Exploratory Search". *ACM Transactions on Information Systems*. 36(4): 1–46. DOI: 10.1145/3231593.

Russell-Rose, T. and T. Tate. (2013). *Designing the Search Experience: The Information Architecture of Discovery*. Newnes.

Sahib, N. G., A. Tombros, and T. Stockman. (2012). "A comparative analysis of the information-seeking behavior of visually impaired and sighted searchers". *Journal of the American Society for Information Science and Technology*. 63(2): 377–391. DOI: 10.1002/asi.21696.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Salminen, J., Y.-H. Liu, S. Şengün, J. M. Santos, S.-g. Jung, B. J. Jansen, S. Sengün, J. M. Santos, S.-g. Jung, and B. J. Jansen. (2020). "The Effect of Numerical and Textual Information on Visual Engagement and Perceptions of AI-Driven Persona Interfaces". In: *Proceedings of the 25th International Conference on Intelligent User Interfaces. IUI '20*. New York, NY, USA: Association for Computing Machinery. 357–368. DOI: 10.1145/3377325.3377492.

Santos, R. L. T., C. Macdonald, and I. Ounis. (2015). "Search Result Diversification". *Foundations and Trends in Information Retrieval*. 9(1): 1–90. DOI: 10.1561/1500000040.

Saracevic, T. (1997). "The stratified model of information retrieval interaction: Extension and applications". In: *Proceedings of the ASIS annual meeting*. Vol. 34. 313–327.

Sarraf, N. (2019). "Mapping the Neural Activities and Affective Dimensions of the ISP Model : Correlates in the Search Exploration , Formulation , and Collection Stages". *PhD thesis*. Queensland University of Technology.

Savage-Knepshield, P. A. and N. J. Belkin. (1999). "Interaction in information retrieval: Trends over time". *Journal of the American Society for Information Science*. 50(12): 1067–1082.

Savenkov, D., P. Braslavski, and M. Lebedev. (2011). "Search snippet evaluation at yandex: lessons learned and future directions". In: *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer. 14–25.

Savolainen, R. (2018). "Information-Seeking Processes as Temporal Developments: Comparison of Stage-based and Cyclic Approaches". *Journal of the Association for Information Science and Technology*. 69(6): 787–797. DOI: 10.1002/asi.24003.

Schaer, P., P. Mayr, and T. Lüke. (2012). "Extending term suggestion with author names". In: *International Conference on Theory and Practice of Digital Libraries*. Springer. 317–322.

Schlötterer, J., C. Seifert, C. Satchell, and M. Granitzer. (2020). "QueryCrumbs search query history visualization – Usability, transparency and long-term usage". *Journal of Computer Languages*. 57: 100941. DOI: 10.1016/j.cola.2020.100941.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Schnabel, T., P. N. Bennett, and T. Joachims. (2019). “Shaping Feedback Data in Recommender Systems with Interventions Based on Information Foraging Theory”. In: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. WSDM ’19*. New York, NY, USA: ACM. 546–554. DOI: 10.1145/3289600.3290974.

Schwartz, M., C. Hash, and L. M. Liebrock. (2010). “Term distribution visualizations with Focus+Context”. *Multimedia Tools and Applications*. 50(3): 509–532. DOI: 10.1007/s11042-010-0479-1.

Setlur, V., E. Hoque, D. H. Kim, and A. X. Chang. (2020). “Sneak Pique: Exploring Autocompletion as a Data Discovery Scaffold for Supporting Visual Analysis”. In: *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology. UIST ’20*. New York, NY, USA: Association for Computing Machinery. 966–978. DOI: 10.1145/3379337.3415813.

Shah, C., J. Pickens, and G. Golovchinsky. (2010). “Role-based results redistribution for collaborative information retrieval”. *Information Processing and Management*. 46(6): 773–781. DOI: 10.1016/j.ipm.2009.10.002.

Shah, C. and R. W. White. (2021). “Task Intelligence for Search and Recommendation”. *Synthesis Lectures on Information Concepts, Retrieval, and Services*. 13(3): 1–160.

Shalini, S., T. Levins, E. L. Robinson, K. Lane, G. Park, and M. Skubic. (2019). “Development and Comparison of Customized Voice-Assistant Systems for Independent Living Older Adults”. In: *Human Aspects of IT for the Aged Population. Social Media, Games and Assistive Environments. HCII 2019. Lecture Notes in Computer Science, vol 11593*. Ed. by J. Zhou and G. Salvendy. Cham: Springer International Publishing. 464–479.

Shao, L., D. Sacha, B. Neldner, M. Stein, and T. Schreck. (2016). “Visual-interactive search for soccer trajectories to identify interesting game situations”. In: *IS and T International Symposium on Electronic Imaging Science and Technology*. Society for Imaging Science and Technology. DOI: 10.2352/issn.2470-1173.2016.1.vda-510.

Sharp, H., J. Preece, and Y. Rogers. (2019). *Interaction Design : Beyond Human-Computer Interaction*. Newark, NJ: John Wiley \& Sons.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Shiga, S., H. Joho, R. Blanco, J. R. Trippas, and M. Sanderson. (2017). "Modelling Information Needs in Collaborative Search Conversations". In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '17*. New York, NY, USA: Association for Computing Machinery. 715–724. DOI: 10.1145/3077136.3080787.

Shneiderman, B. (1983). "Direct Manipulation: A Step Beyond Programming Languages". *Computer*. 16(8): 57–69. DOI: 10.1109/MC.1983.1654471.

Shneiderman, B., C. Plaisant, M. S. Cohen, S. Jacobs, N. Elmqvist, and N. Diakopoulos. (1997). *Designing the User Interface : Strategies for Effective Human-Computer Interaction*. 6 Edition. Addison Wesley.

Shokouhi, M. (2013). "Learning to personalize query auto-completion". In: *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '13*. 103. DOI: 10.1145/2484028.2484076.

Silva, N., T. Blascheck, R. Jianu, N. Rodrigues, D. Weiskopf, M. Raubal, and T. Schreck. (2019). "Eye Tracking Support for Visual Analytics Systems: Foundations, Current Applications, and Research Challenges". In: *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications. ETRA '19*. New York, NY, USA: ACM. 11:1–11:10. DOI: 10.1145/3314111.3319919.

Skarbez, R., F. P. Brooks Jr., and M. C. Whitton. (2018). "Immersion and Coherence in a Stressful Virtual Environment". In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology. VRST '18*. New York, NY, USA: ACM. 24:1–24:11. DOI: 10.1145/3281505.3281530.

Skarlatidou, A., A. Hamilton, M. Vitos, and M. Haklay. (2019). "What do volunteers want from citizen science technologies? A systematic literature review and best practice guidelines". *Journal of Science Communication*. 18(01). DOI: 10.22323/2.18010202.

Skov, M., B. Larsen, and P. Ingwersen. (2008). "Inter and intra-document contexts applied in polyrepresentation for best match IR". *Information Processing & Management*. 44(5): 1673–1683.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Slater, M., B. Spanlang, and D. Corominas. (2010). “Simulating virtual environments within virtual environments as the basis for a psychophysics of presence”. In: *ACM SIGGRAPH 2010 Papers, SIGGRAPH 2010*. Vol. 29. No. 4. Association for Computing Machinery, Inc. 1–9. DOI: 10.1145/1778765.1778829.

Smith, C. L., J. Gwizdka, and H. Feild. (2016). “Exploring the Use of Query Auto Completion: Search Behavior and Query Entry Profiles”. In: *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval. CHIIR ’16*. New York, NY, USA: Association for Computing Machinery. 101–110. DOI: 10.1145/2854946.2854975.

Smith, C. L., J. Gwizdka, and H. Feild. (2017). “The use of query auto-completion over the course of search sessions with multifaceted information needs”. *Information Processing and Management*. 53(5): 1139–1155.

Smith, C. L. and S. Y. Rieh. (2019). “Knowledge-Context in Search Systems: Toward Information-Literate Actions”. In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. ACM. 55–62.

Soleymani, M., M. Riegler, and P. Halvorsen. (2017). “Multimodal analysis of image search intent: Intent recognition in image search from user behavior and visual content”. In: *ICMR 2017 - Proceedings of the 2017 ACM International Conference on Multimedia Retrieval*. New York, NY, USA: Association for Computing Machinery, Inc. 251–259. DOI: 10.1145/3078971.3078995.

Sonnenwald, D., M. Bates, S. Lin Chang, C. Kuhlthau, G. Olson, J. Olson, M. K. Buckland, K. Järvelin, T. Saracevic, M. Thelwall, J. M. Carrroll, B. Nardi, A. Dillon, J. McGann, H. S. Crew, D. Bawden, and J. Meadows. (2016). *Theory Development in the Information Sciences*. University of Texas Press. DOI: 10.7560/308240.

Sparck Jones, K. (1994). “Towards Better NLP System Evaluation”. In: *Proceedings of the Workshop on Human Language Technology. HLT ’94*. Stroudsburg, PA, USA: Association for Computational Linguistics. 102–107. DOI: 10.3115/1075812.1075833.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Spiller, M., Y.-H. Liu, M. Z. Hossain, T. Gedeon, J. Geissler, and A. Nürnberger. (2021). "Predicting Visual Search Task Success from Eye Gaze Data as a Basis for User-Adaptive Information Visualization Systems". *ACM Transactions on Intelligent Systems and Technology*. 11(2). DOI: 10.1145/3446638.

Spink, A., H. Greisdorf, and J. Bateman. (1998). "From highly relevant to not relevant: examining different regions of relevance". *Information Processing & Management*. 34(5): 599–621.

Stasko, J., C. Görg, and Z. Liu. (2008). "Jigsaw: Supporting Investigative Analysis through Interactive Visualization". *Information Visualization*. 7(2): 118–132. DOI: 10.1057/palgrave.ivs.9500180.

Stebbins, R. A. (2009). "Leisure and Its Relationship to Library and: Information Science: Bridging the Gap". *Library Trends*. 57(4): 618–631. DOI: 10.1353/lib.0.0064.

Stebbins, R. A. (2007). *Serious leisure : a perspective for our time*. Transaction Publishers. 156.

Steichen, B., C. Conati, and G. Carenini. (2014). "Inferring Visualization Task Properties, User Performance, and User Cognitive Abilities from Eye Gaze Data". *ACM Transactions on Interactive Intelligent Systems*. 4(2): 1–29. DOI: 10.1145/2633043.

Tamine, L. and M. Daoud. (2018). "Evaluation in Contextual Information Retrieval: Foundations and Recent Advances Within the Challenges of Context Dynamicity and Data Privacy". *ACM Comput. Surv.* 51(4): 78:1–78:36. DOI: 10.1145/3204940.

Tang, R. and Y. Song. (2018). "Cognitive styles and eye movement patterns: an empirical investigation into user interactions with interface elements and visualisation objects of a scientific information system". *Information Research*. 23(2).

Taramigkou, M., D. Apostolou, and G. Mentzas. (2017). "Supporting Creativity through the Interactive Exploratory Search Paradigm". *International Journal of Human–Computer Interaction*. 33(2): 94–114. DOI: 10.1080/10447318.2016.1220104.

Taylor, R. S. (1967). "Question-negotiation and information seeking in libraries". *College and Research Libraries*. 29(3): 178–194.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Thomas, P., M. Czerwinski, D. McDuff, N. Craswell, and G. Mark. (2018). “Style and Alignment in Information-Seeking Conversation”. *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - CHIIR ’18*: 42–51. DOI: 10.1145/3176349.3176388.

Toker, D., C. Conati, and G. Carenini. (2018). “User-adaptive Support for Processing Magazine Style Narrative Visualizations”. *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - IUI ’18*: 199–204. DOI: 10.1145/3172944.3173009.

Toker, D., S. Lallé, and C. Conati. (2017). “Pupillometry and Head Distance to the Screen to Predict Skill Acquisition During Information Visualization Tasks”. In: *Proceedings of the 22nd International Conference on Intelligent User Interfaces - IUI ’17*. 221–231. DOI: 10.1145/3025171.3025187.

Toker, D., R. Moro, J. Simko, M. Bielikova, and C. Conati. (2019). “Impact of English Reading Comprehension Abilities on Processing Magazine Style Narrative Visualizations and Implications for Personalization”. In: *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization. UMAP ’19*. New York, NY, USA: ACM. 309–317. DOI: 10.1145/3320435.3320447.

Toms, E., T. MacKenzie, C. Jordan, H. O’Brien, L. Freund, S. Toze, E. Dawe, and A. MacNutt. (2007). “How task affects information search”. In: *Workshop Pre-proceedings in Initiative for the Evaluation of XML Retrieval (INEX)*. 337–341.

Toms, E. G. (2011). “task-based information searching and retrieval elaine toms interactive information seeking behaviour and retrieval”. In: *Interactive Information Seeking, Behaviour and Retrieval*. Ed. by I. Ruthven and D. Kelly. Facet Publishing. 43–60.

Tou, F. N., M. D. Williams, R. Fikes, D. A. Henderson Jr, and T. W. Malone. (1982). “RABBIT: An Intelligent Database Assistant.” In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 314–318.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Trippas, J. R., D. Spina, F. Scholer, A. H. Awadallah, P. Bailey, P. N. Bennett, R. W. White, J. Liono, Y. Ren, F. D. Salim, and M. Sanderson. (2019). "Learning About Work Tasks to Inform Intelligent Assistant Design". In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*. *CHIIR '19*. New York, NY, USA: ACM. 5–14. DOI: 10.1145/3295750.3298934.

Trippas, J. R., D. Spina, L. Cavedon, H. Joho, and M. Sanderson. (2018). "Informing the Design of Spoken Conversational Search". *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - CHIIR '18*: 32–41. DOI: 10.1145/3176349.3176387.

Tunkelang, D. (2009). *Faceted search*. Vol. 1. No. 1. Morgan & Claypool Publishers. 1–80. DOI: 10.2200/s00190ed1v01y200904icr005.

Turpin, A., Y. Tsegay, D. Hawking, and H. E. Williams. (2007). "Fast generation of result snippets in web search". In: *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. 127–134.

Urgo, K., J. Arguello, and R. Capra. (2019). "Anderson and Krathwohl's Two-Dimensional Taxonomy Applied to Task Creation and Learning Assessment". In: *Proceedings of the 5th ACM SIGIR International Conference on the Theory of Information Retrieval*. DOI: 10.1145/3341981.3344226.

Vakkari, P. and S. Huuskonen. (2012). "Search effort degrades search output but improves task outcome". 63: 657–670.

Vakkari, P. (2000a). "Cognition and changes of search terms and tactics during task performance: A longitudinal case study". In: *Content-Based Multimedia Information Access-Volume 1*. Le Centre de Hautes Etudes Internationales D'informatique Documentaire. 894–907.

Vakkari, P. (2000b). "Relevance and contributing information types of searched documents in task performance". In: *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM. 2–9.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Vakkari, P. (2020). “The usefulness of search results: A systematization of types and predictors”. In: *CHIIR 2020 - Proceedings of the 2020 Conference on Human Information Interaction and Retrieval. CHIIR ’20*. New York, NY, USA: Association for Computing Machinery. 243–252. DOI: 10.1145/3343413.3377955.

Vakkari, P. and N. Hakala. (2000). “Changes in relevance criteria and problem stages in task performance”. *Journal of documentation*. 56(5): 540–562.

Vakkari, P. and M. Pennanen. (2001). “Sources, relevance and contributory information of documents in writing a research proposal: a longitudinal case study”. *The new review of information behaviour research*. 2(November): 217–232.

Vakkari, P., M. Völske, M. Potthast, M. Hagen, and B. Stein. (2018). “Predicting Retrieval Success Based on Information Use for Writing Tasks”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Ed. by E. Méndez, F. Crestani, C. Ribeiro, G. David, and J. C. Lopes. Vol. 11057 LNCS. Cham: Springer International Publishing. 161–173. DOI: 10.1007/978-3-030-00066-0{_}14.

Vakulenko, S., K. Revoredo, C. Di Ciccio, and M. M. de Rijke. (2019). “QRFA: A Data-Driven Model of Information-Seeking Dialogues”. In: *Advances in Information Retrieval. ECIR 2019. Lecture Notes in Computer Science, vol 11437*. Ed. by L. Azzopardi, B. Stein, N. Fuhr, P. Mayr, C. Hauff, and D. Hiemstra. Cham: Springer International Publishing. 541–557. DOI: 10.1108/jd.2012.27868eaa.002.

Vinchira, A. V. (2019). “Modelling the information practices of music fans living in Medellín, Colombia”. *Information Research*. 24(3). URL: http://informationr.net/ir/24-3/paper833.html.

Vtyurina, A., A. Fourney, M. R. Morris, L. Findlater, and R. W. White. (2019). “Verse: Bridging screen readers and voice assistants for enhanced eyes-free web search”. In: *ASSETS 2019 - 21st International ACM SIGACCESS Conference on Computers and Accessibility*. Association for Computing Machinery, Inc. 414–426. DOI: 10.1145/3308561.3353773.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Vtyurina, A., D. Savenkov, E. Agichtein, and C. L. A. Clarke. (2017). "Exploring Conversational Search With Humans, Assistants, and Wizards". In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. CHI EA '17*. New York, NY, USA: ACM. 2187–2193. DOI: 10.1145/3027063.3053175.

Wacholder, N. and L. Liu. (2008). "Assessing term effectiveness in the interactive information access process". *Information Processing and Management*. 44(3): 1022–1031. DOI: 10.1016/j.ipm.2007.07.011.

Wacholder, N., L. Liu, and Y.-H. Liu. (2006). "Selecting books: A performance-based study". *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*: 337. DOI: 10.1145/1141753.1141833.

Wagner, N., K. Hassanein, and M. Head. (2014). "The impact of age on website usability". *Computers in Human Behavior*. 37(Aug.): 270–282. DOI: 10.1016/J.CHB.2014.05.003.

Walker, D. E. (1971). "Interactive Bibliographic Search: The User/Computer Interface". In: *"The User Interface for Interactive Search of Bibliographic Data Bases*. Ed. by D. E. Walker. Montvale, New Jersey: AFIPS Press. 404.

Walsh, D. and M. Hall. (2015). "Just looking around: Supporting casual users initial encounters with Digital Cultural Heritage". In: *CEUR Workshop Proceedings*. Vol. 1338.

Walsh, J. A., P. J. Cobb, W. de Fremery, K. Golub, H. Keah, J. Kim, J. Kiplang'at, Y.-H. Liu, S. Mahony, S. G. Oh, C. A. Sula, T. Underwood, and X. Wang. (2021). "Digital humanities in the iSchool". *Journal of the Association for Information Science and Technology*. n/a(n/a). DOI: 10.1002/asi.24535.

Ward, A. R. and R. Capra. (2020). "Immersive Search: Using Virtual Reality to Examine How a Third Dimension Impacts the Searching Process". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY, USA: Association for Computing Machinery (ACM). 1621–1624. DOI: 10.1145/3397271.3401303.

Weiskopf, D. (2019). "Vis4Vis: Visualization for (Empirical) Visualization Research". 18041: 1–16. URL: http://arxiv.org/abs/1908.00611.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

White, R. W. (2016). *Interactions with search systems*. Cambridge University Press, New York.

White, R. W. and R. A. Roth. (2009). "Exploratory search: Beyond the query-response paradigm". *Synthesis lectures on information concepts, retrieval, and services*. 1(1): 1–98.

White, R. W. (2018). "Opportunities and challenges in search interaction". *Communications of the ACM*. 61(12): 36–38. DOI: 10.1145/3195180.

White, R. W., I. Ruthven, J. M. Jose, and C. J. V. Rijsbergen. (2005). "Evaluating implicit feedback models using searcher simulations". *ACM Transactions on Information Systems*. 23(3): 325–361. DOI: 10.1145/1080343.1080347.

Whiting, S. and J. M. Jose. (2014). "Recent and robust query autocompletion". In: *Proceedings of the 23rd international conference on World wide web - WWW '14*. New York, New York, USA: ACM Press. 971–982. DOI: 10.1145/2566486.2568009.

WHO. (2011). *World Report on Disability*. 325. URL: https://www.who.int/publications/i/item/9789241564182.

WHO. (2019). "Global Health Estimates: Life expectancy and leading causes of death and disablity". URL: https://www.who.int/docs/default-source/gho-documents/global-health-estimates/ghe2019_life-table-methods.pdf.

WHO. (2021). *World Health Statistics 2021*. URL: https://apps.who.int/iris/bitstream/handle/10665/342703/9789240027053-eng.pdf.

Williamson, K. (2018). "Research concepts". In: *Research methods: Information, Systems, and Contexts*. Ed. by K. Williamson and G. Johanson. 2nd ed. Chandos Publishing. Chap. 1. 3–25. DOI: 10.1016/B978-0-08-102220-7.00001-7.

Wilson, M., J. Hurlock, and M. Wilson. (2012a). "Keyword clouds". In: *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12*. New York, New York, USA: ACM Press. 2069–2074. DOI: 10.1145/2212776.2223754.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Wilson, M. J. and M. L. Wilson. (2013). "A comparison of techniques for measuring sensemaking and learning within participant-generated summaries". *Journal of the American Society for Information Science and Technology*. 64(2): 291–306. DOI: 10.1002/asi.22758.

Wilson, M. L. (2011). *Search User Interface Design*. Vol. 3. No. 3. Morgan & Claypool Publishers. 1–143. DOI: 10.2200/S00371ED1V01Y201111ICR020.

Wilson, M. L., B. Alhodaithi, and M. Hurst. (2012b). "A Diary Study of Information Needs Produced in Casual-Leisure Reading Situations". In: *Searching4Fun Workshop at ECIR 2012*. URL: http://cs.swan.ac.uk/~csmax/pubs/search4fun-reading.pdf.

Wilson, M. L., M. C. Schraefel, and R. W. White. (2009). "Evaluating advanced search interfaces using established information-seeking models". *Journal of the American Society for Information Science and Technology*. 60(7): 1407–1422.

Wilson, M. L. (2017). "The Tetris Model of Resolving Information Needs within the Information Seeking Process". In: *CHIIR '17*. 147–154.

Wilson, T. D. (1999). "Models in information behaviour research". *Journal of Documentation*. 55(3): 249–270. DOI: 10.1108/EUM0000000007145.

Wilson, T. D. (2016). "A general theory of human information behaviour". *Information Research*. 21(4). URL: http://informationr.net/ir/21-4/isic/isic1601.html.

Winther, K. D., S. Livingstone, and M. Saeed. (2019). "Growing up in a connected world, Innocenti Research Report". *Tech. rep.* UNICEF. URL: https://www.unicef-irc.org/publications/1060-growing-up-in-a-connected-world.html.

Wittek, P., Y.-H. Liu, S. Darányi, T. Gedeon, and I. S. Lim. (2016). "Risk and ambiguity in information seeking: Eye gaze patterns reveal contextual behavior in dealing with uncertainty". *Frontiers in Psychology*. 7(NOV): 1790. DOI: 10.3389/fpsyg.2016.01790.

Wu, D. and S. Liang. (2018). *Mobile Search Behaviors: An In-depth Analysis Based on Contexts, APPs, and Devices*. Vol. 10. No. 2. Morgan & Claypool Publishers. i–159. DOI: 10.2200/S00831ED1V01Y201802ICR063.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Wu, K. C. and T. Y. Hsieh. (2016). "Affective choosing of clustering and categorization representations in e-book interfaces". *Aslib Journal of Information Management*. 68(3): 265–285. DOI: 10.1108/AJIM-12-2015-0191.

Wu, M.-M. (2005). "Understanding patrons' micro-level information seeking (MLIS) in information retrieval situations". *Information Processing & Management*. 41(4): 929–947. DOI: 10.1016/j.ipm.2004.08.007.

Wu, M.-M. and Y.-H. Liu. (2003). "Intermediary's information seeking, inquiring minds, and elicitation styles". *Journal of the American Society for Information Science and Technology*. 54(12): 1117–1133. DOI: 10.1002/asi.10323.

Wu, M.-M. and Y.-H. Liu. (2011). "On intermediaries' inquiring minds, elicitation styles, and user satisfaction". *Journal of the American Society for Information Science and Technology*. 62(12): 2396–2403. DOI: 10.1002/asi.21644.

Wu, W.-C. and D. Kelly. (2014). "Online search stopping behaviors: An investigation of query abandonment and task stopping". *Proceedings of the American Society for Information Science and Technology*. 51(1): 1–10. DOI: 10.1002/meet.2014.14505101030.

Wu, W.-C., D. Kelly, A. Edwards, and J. Arguello. (2012). "Grannies, tanning beds, tattoos and NASCAR: Evaluation of search tasks with varying levels of cognitive complexity". In: *Proceedings of the 4th Information Interaction in Context Symposium*. ACM. 254–257.

Wu, Y., Y. Liu, Y.-H. R. Tsai, and S.-T. Yau. (2019). "Investigating the role of eye movements and physiological signals in search satisfaction prediction using geometric analysis". *Journal of the Association for Information Science and Technology*. 70(9): 981–999. DOI: 10.1002/asi.24240.

Xie, H. (2000). "Shifts of interactive intentions and information-seeking strategies in interactive information retrieval". *Journal of the American Society for information Science*. 51(9): 841–857.

Xie, I. (2009). "Dimensions of tasks: influences on information-seeking and retrieving process". *Journal of Documentation*. 65(3): 339–366.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Ye, Y., F. Sauer, K.-L. Ma, K. Aditya, and J. Chen. (2020). "A User-centered Design Study in Scientific Visualization Targeting Domain Experts". *IEEE Transactions on Visualization and Computer Graphics*. 2626(c): 1–1. DOI: 10.1109/tvcg.2020.2970525.

Yigitbas, E., A. Hottung, S. M. Rojas, A. Anjorin, S. Sauer, and G. Engels. (2019). "Context- and Data-Driven Satisfaction Analysis of User Interface Adaptations Based on Instant User Feedback". *Proc. ACM Hum.-Comput. Interact.* 3(EICS). DOI: 10.1145/3331161.

Yoon, K., R. Dols, L. Hulscher, and T. Newberry. (2016). "An exploratory study of library website accessibility for visually impaired users". *Library & Information Science Research*. 38(3): 250–258. DOI: 10.1016/J.LISR.2016.08.006.

Yuan, X. J., N. Belkin, and J. Y. Kim. (2002). "The relationship between ASK and relevance criteria". In: *SIGIR Forum (ACM Special Interest Group on Information Retrieval)*. Association for Computing Machinery (ACM). 359–360.

Yuan, X. and N. J. Belkin. (2007). "Supporting Multiple Information-seeking Strategies in a Single System Framework". In: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '07*. New York, NY, USA: ACM. 247–254. DOI: 10.1145/1277741.1277786.

Yue, Z., S. Han, and D. He. (2014). "Modeling search processes using hidden states in collaborative exploratory web search". In: *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. 820–830. DOI: 10.1145/2531602.2531658.

Zhang, A., A. Goyal, W. Kong, H. Deng, A. Dong, Y. Chang, C. A. Gunter, and J. Han. (2015). "adaQAC: : Adaptive Query Auto-Completion via Implicit Negative Feedback". In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '15*. 143–152. DOI: 10.1145/2766462.2767697.

Zhang, F., K. Zhou, Y. Shao, C. Luo, M. Zhang, and S. Ma. (2018). "How Well do Offline and Online Evaluation Metrics Measure User Satisfaction in Web Image Search?" In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM. 615–624.

The version of record is available at: http://dx.doi.org/10.1561/1500000073

Zhang, H. (2018). "Beyond Query-Oriented Highlighting: Investigating the Effect of Snippet Text Highlighting in Search User Behavior". *Computational Intelligence and Neuroscience*. Dec.: 1–12. DOI: 10.1155/2018/7836969.

Zhang, Y. and C. Liu. (2020). "Users' Knowledge Use and Change during Information Searching Process: A Perspective of Vocabulary Usage". In: *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. New York, NY, USA: ACM. 47–56. DOI: 10.1145/3383583.3398532.

Zhang, Y. and R. Capra. (2019). "Understanding How People Use Search to Support Their Everyday Creative Tasks". In: *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval. CHIIR '19*. New York, NY, USA: ACM. 153–162. DOI: 10.1145/3295750.3298936.